الأساليب الإحصائية

الوصفية والاستدلالية في تحليل البيانات

د.عبد الله عامر الهمالي



الأساليب الإحصائية الوصفية والاستدلالية

في تحليل البيانات

فهرسة أثناء النشر إعداد إدارة الشئون الفنية – دار الكتب المصرية

الهمالي ، عبد الله عامر

الأساليب الإحصائية الوصفية والاستدلالية في تحليل البيانات/

تأليف: عبد الله عامر الهمالي

ط1ـ القاهرة: المجموعة العربية للتدريب والنشر، 2013.

665 ص: 17 × 24 سم.

الترقيم الدولي: 978-977-722-050-7

1- الاجتماع، علم الطرق الإحصائية

2- الإحصاء التحليلي

أـ العنوان

ديوي: 301,0182 رقم الإيداع: 2014/3637

الطبعة الأولى
2014



الناشر

المجموعة العربية للتدريب والنشر

8أ شارع أحمد فخري – مدينة نصر – القاهرة – مصر

تليفاكس: 22759945 - 22739110 (00202)

الموقع الإلكتروني: www.arabgroup.net.eg

E-mail: info@arabgroup.net.eg

elarabgroup@yahoo.com

الأساليب الإحصائية الوصفية والاستدلالية في تحليل البيانات

تأليف

د. عبد اللّه عامر الهمالي

أستاذ الإحصاء الاجتماعي ومناهج البحث

كلية الآداب- جامعة بنغازي

الناشر

المجموعة العربية للتدريب والنشر



2014

بسم الله الرحمن الرحيم

﴿ وَمِنَ ٱلنَّاسِ وَٱلدَّوَآبِّ وَٱلۡأَنۡعَٰمِ مُخۡتَلِفٌ أَلۡوَٰنُهُۥ كَذَٰلِكَۗ إِنَّمَا يَخۡشَى ٱللَّهَ
مِنۡ عِبَادِهِ ٱلۡعُلَمَٰٓؤُاْۗ إِنَّ ٱللَّهَ عَزِيزٌ غَفُورٌ ﴿٢٨﴾ ﴾

صدق الله العظيم

(سورة فاطر، الآية 28)

الإهداء

إلى أعز الأحفاد ..

أحمد

يوسف

وعبد اللّه

بداية الحاضر .. وكل المستقبل

المحتويات مُخْتصرةَ

تقديم 25

الجزء الأول: الإحصاءات الوصفية: الأحادية 29

- الفصل الأول: الإحصاء والبيانات وتصميم البحث 31
- الفصل الثاني: المتغيرات والقياس 47
- الفصل الثالث: تحليل البيانات التكرارية 73
- الفصل الرابع: مقاييس النزعة المركزية ومقاييس التشتت 97
- الفصل الخامس: المنحنى الطبيعي 137

الجزء الثاني: الإحصاءات الوصفية الثنائية 163

- الفصل السادس: تحليل جداول التقاطع: استقصاء العلاقة بين متغيرين: الوصف الجدولي لبيانات مقاسة على المستويين الاسمي والترتيبي 165
- الفصل السابع: الوصف العددي للبيانات الاسمية: مقاييس التطابق الثنائية 187
- الفصل الثامن: التطابق بين متغيرات تم قياسها على المستوى الترتيبي الوصف العددي للبيانات الترتيبية 213
- الفصل التاسع: التطابق بين متغيرات تم قياسها على المستوى ذي المسافات والنسبي: شكل الانتشار وخط الانحدار 243

الجزء الثالث: الإحصاءات الاستدلالية البارامترية (حالة العينة الواحدة) 285

- الفصل العاشر: توزيعات المعاينة - التقدير، وفترات الثقة 287
- الفصل الحادي عشر: اختبار الفروض: اختبار Z لمتوسط عينة واحدة 321
- الفصل الثاني عشر: اختبار t لمتوسط حسابي لعينة واحدة 359

الجزء الرابع: الإحصاءات الاستدلالية البارامترية لعينتين مستقلتين أو أكثر 377

- الفصل الثالث عشر: اختبار t لعينتين ذات وسطين حسابيين متساويين 379
- الفصل الرابع عشر: فروق الدلالة: اختبار F لأكثر من وسطين متساويين: تحليل التباين 415

الجزء الخامس: الإحصاءات الوصفية المتعددة 437

- الفصل الخامس عشر: التوسع في جداول التقاطع: إضافة متغيرات التحكم 439
- الفصل السادس عشر: الارتباط الجزئي والانحدار المتعدد 461

الجزء السادس: الإحصاءات اللا برامترية 443

- الفصل السابع عشر: اختبار مربع كاي (X^2) لفروق الدلالة لعينة واحدة لتوزيع تكراري 445
- الفصل الثامن عشر: اختبار مربع كاي (X^2) للاستقلال 463
- الفصل التاسع عشر: اختبار توزيع ذي الحدين لعينة واحدة 507
- الفصل: العشرون: الأساليب الإحصائية للبيانات الترتيبية: اختبارات: مان – وتني، ولكوكسن، كروسكال – وليز، وفريدمان 527

الملاحــق 617

المحتويات مُفصلةً

مقدمة 25

الجزء الأول: الإحصاءات الوصفية: الأحادية 29

الفصل الأول: الإحصـاء والبيانـات وتصميـم البحث 31

- مقدمة 31

1- عملية البحث الاجتماعي 34

- تحديد أهداف البحث 34
- مراجعة الأدبيات المتعلقة بالمشكلة المطروحة 34
- صياغة الفروض 34
- القياس 35
- طرق جمع البيانات 35
- تحليل البيانات 35

2- الأسس المتعلقة بتصميم البحث 36

- المجتمع، عناصر المجتمع، ووحدات التحليل 36
- العينات وإطار المعاينة 37
- المتغيرات 39
- الفئات، والقيم، والبيانات 39
- المتغيرات المستقلة والمتغيرات التابعة 41
- منطق العلاقة بين المتغيرات 42
- متغيرات التحكم 44

- الخلاصة 44
- أسئلة للمراجعة 45
- الهوامش والمصادر 46

الفصل الثاني: المتغيرات والقياس 47

التعريفات التصورية والإجرائية للمتغيرات 49
• الإطار النظري 50
• الأجندة المحددة مسبقاً للبحث 50
• حب الاستطلاع يقود للبحث 51
1- تصنيف المتغيرات وفقاً لمستوى القياس 57
• المقياس الاسمي 57
• المقياس الترتيبي 59
• مقياس ذو المسافات والنسبي 61
2- الفرق بين المتغيرات المقولية والمتغيرات العددية 62
• المتغيرات المقولية: الفرق بين المتغيرات الاسمية والترتيبية 63
• المتغيرات العددية 64
3- المتغيرات المنفصلة والمتغيرات المتصلة 65
- خلاصة 68
- أسئلة للمراجعة 69
- الهوامش والمصادر 72

الفصل الثالث: تحليل البيانات التكرارية 73

- مقدمة 73
1- أنواع الإحصاءات الوصفية 74
2- الجداول التكرارية 76
3- أشكال التوزيعات التكرارية 77

- جداول البيانات المدرجة 77
- جداول التوزيعات التكرارية البسيطة 79
- جداول التكرارات النسبيِّة 81
- جداول التكرار المتجمع 86

4- إجراءات توليد التكرارات باستخدام SPSS 92

– أسئلة للمراجعة 94

– الهوامش والمصادر 95

الفصل الرابع: مقاييس النـزعة المركزية ومقاييس التشتت 97

أولاً: مقاييس النزعة المركزية 97

1- المنوال 99

2- الوسيط 101

3- المتوسط 103

4- الخصائص العامة للمتوسط والوسيط والمنوال 106

5- الخصائص المؤثرة في التفسير 106

- تأثير القيم المتطرفة 106
- الخصائص المؤثرة في الاستدلالات 109
- الثبات في سحب العينات العشوائية 109
- مجموع الانحرافات عن المتوسط الحسابي تساوي صفراً (0) 113

6- اختيار مقاييس النزعة المركزية 114

7- استخدام مقاييس النزعة المركزية 115

ثانياً: مقاييس التشتت 116

1- المدى 118

2- المدى الربيعي 118

3- الانحراف المعياري 120

4- معامل التباين النسبي (CRV) 123

5- مؤشر التباين النوعي (IQV) 124
6- استخدام مقاييس التشتت 130
7- إجراءات حساب مقاييس النزعة المركزية ومقاييس التشتت باستخدام SPSS 132
- أسئلة للمراجعة 133
- الهوامش والمصادر 135

الفصل الخامس: المنحنى الطبيعي 137
- مقدمة 137
1- التوزيع الطبيعي 137
2- المنحنى الطبيعي كأداة مساعدة لوصف البيانات 143
3- المنحنى الطبيعي كأداة للإحصاء الاستدلالي 143
4- استخدام المنحنيات الطبيعية لوصف توزيع 143
5- درجات Z 148
6- المساحة بين المتوسط الحسابي ونقطة على التوزيع 150
7- المساحة أبعد من نقطة على التوزيع 152
8- المساحة بين نقطتين على التوزيع الطبيعي 154
9- حساب القيم من درجات Z 155
- أسئلة للمراجعة 159
- الهوامش والمصادر 162

الجزء الثاني: الإحصاءات الوصفية الثنائية 163

الفصل السادس: تحليل جداول التقاطع: استقصاء العلاقة بين متغيرين
الوصف الجدولي لبيانات مقاسة على المستويين الاسمي والترتيبي 165
- مقدمة 165
1- العلاقة بين المتغيرات 166
2- نمذجة العلاقة بين متغيرين 168

- علاقة مباشرة أحادية الاتجاه "مكان الإقامة كمتغير تابع" 169
- علاقة أحادية الاتجاه "مكان الإقامة كمتغير تابع" 169
- علاقة مباشرة ثنائية الاتجاه بين مكان الإقامة والدخل بشكل متعمد تبادلياً 169

3- وصف البيانات المقولية باستخدام جداول التقاطع 171
4- جداول التقاطع ذات التوزيعات النسبية 175
5- إجراءات توليد جداول التقاطع CROSS TABLES باستخدام SPSS 177
6- تفسير جداول التوافق: نمط وقوة العلاقة 179
7- تفسير الجداول المتقاطعة عندما يكون لدينا متغيران مقاسان على المستوى الترتيبي 180

- اتجاه العلاقة 180
- علاقة الاتساق 182

– الخلاصة 184
– أسئلة للمراجعة 184
– الهوامش والمصادر 186

الفصل السابع: الوصف العددي للبيانات الاسمية: مقاييس التطابق 187

– مقدمة 187

- مقاييس التطابق كإحصاءات وصفية 189
- مقاييس التطابق للمتغيرات الاسمية 192
- منطقة نسبة التخفيض في الخطأ (PRE) 195
- خصائص لامبيدا 199
- إجراءات توليد لامبيدا LAMBDA باستخدام SPSS 202
- حدود استخدام لامبيدا 203

– أسئلة للمراجعة 210
– الهوامش والمصادر 212

الفصل الثامن: التطابق بين متغيرات تم قياسها على المستوى الترتيبي

الوصف العددي للبيانات الترتيبية 213
- مقدمة 213
• الأزواج المتوافقة 214
• الأزواج غير المتوافقة 218
• مقاييس التطابق للمتغيرات الترتيبية 219
• جاما GAMMA 221
• سومرز SOMERŚ D D 225
• كندل – تاو B 227
• كندل تاو – C 228
• إجراء مقاييس التطابق باستخدام SPSS 229
• تفسير مقاييس التطابق من مخرجات SPSS 230
• معامل سبيرمان للرتب 231
• رتب الدرجات المتعادلة 235
• المعادلة الخاصة بارتباط سبيرمان 236
- الخلاصة 238
- أسئلة للمراجعة 239
- الهوامش والمصادر 242

الفصل التاسع: التطابق بين متغيرات تم قياسها على المستوى ذي المسافات

والنسبي: شكل الانتشار وخط الانحدار 243
- مقدمة 243
• الرسم الانتشاري 244
• خط الانحدار 246
• معامل ارتباط بيرسون (r) 254
• تفسير التباين: معامل التحديد أو التقدير 256

- التباين الكلي: المفسر وغير المفسر 261
- القياس والمتغيرات الديمومية 263
- إجراءات الرسم البياني، والارتباط والانحدار باستخدام SPSS 266
- إجراءات إضافة خط الانحدار للرسم البياني باستخدام SPSS 267
- إجراءات الانحدار 269
- تفسير مخرجات Spss للانحدار 270
- اختبار t لمعامل الارتباط 271
- اختبار الدلالة لارتباط بيرسون (r) باستخدام SPSS 276
- تفسير اختبار الدلالة لمعامل ارتباط بيرسون (r) من خلال مخرجات SPSS 277
- اختبار t لمعامل ارتباط سبيرمان rho 277

– أسئلة للمراجعة 281
– الهوامش والمصادر 284

الجزء الثالث: الإحصاءات الاستدلالية البارامترية 285

الفصل العاشر: توزيعات المعاينة ـ التقدير، وفترات الثقة 287

أولاً: توزيعات المعاينة 287
- العينات العشوائية 291
- المعاينة العشوائية الطبقية 291
- توزيع المعاينة لإحصاء العينة 293
- نظرية تقارب التوزيعات الاحتمالية 298

ثانياً: التقدير وفترات الثقة 299
- التقدير 301
- تغير مستوى الثقة 306
- إجراء توليد عينات عشوائية مكررة باستخدام برنامج SPSS 309
- حساب متوسط العينة 310

- إجراء تكرار المعاينة 310
- إجراء فترات الثقة 311
- تغير حجم العينة 312
- اختيار حجم العينة 314

– أسئلة للمراجعة 316
– الهوامش والمصادر 319

الفصل الحادي عشر: اختبار الفروض: اختبار Z لمتوسط عينة واحدة 321

– اختبار الفرض: فكرة عامة 321
– نموذج الخطوات الخمس لاختبار الفرض 327

الخطوة الأولى: صياغة الفرض الصفري والفرض البديل 327

الخطوة الثانية: اختيار اختبار الدلالة 330

الخطوة الثالثة: حساب درجة العينة 332

الخطوة الرابعة: بيان الدرجة الحرجة أو المنطقة الحرجة 333

الخطوة الخامسة: اتخاذ القرار: مقارنة درجة العينة بالدرجة الحرجة 340

– ماذا يعني قرار العجز في رفض الفرض الصفري؟ 343
– ماذا يعني قرار رفض الفرض الصفري؟ 344
– اختبار ثنائي الجانب لـِ Z لمتوسط مفرد 345
– اختبار Z أحادي الجانب لمتوسط مفرد 349
– المحاذير المتعلقة باختبار الفرض: قياس حجم التأثير 352
– قياس حجم التأثير 355
– أسئلة للمراجعة 356
– الهوامش والمصادر 358

الفصل الثاني عشر: اختبار t لمتوسط حسابي لعينة واحدة 359

– مقدمة 359

- • توزيع T 360
- • اختبار T لمتوسط حسابي لعينة واحدة 362
- • مثال (1): اختبار أحادي الجانب 365
- • مثال (2): اختبار ثنائي الجانب 368
- • إجراء توليد اختبار T لعينة واحدة باستخدام SPSS: 370
- الخلاصة 371
- أسئلة للمراجعة 373
- الهوامش والمصادر 375

الجزء الرابع: الإحصاءات الاستدلالية البارامترية لعينتين مستقلتين أو أكثر 377

الفصل الثالث عشر: اختبار t لعينتين ذات وسطين حسابيين متساويين 379

- مقدمة 379
- المتغيرات المستقلة والمتغيرات التابعة 382
- توزيعات المعاينة للفرق بين وسطين 383
- اختبار T لعينتين مستقلتين ذات وسطين متساويين 386
- إجراء اختبار T لعينتين مستقلتين باستخدام برنامج SPSS 391
- تفسيرات مخرجات اختبار T 392
- حساب حجم التأثير لاختبار العينات المستقلة 394
- اختبار أحادي الجانب وثنائي الجانب 395
- العينات المستقلة والعينات التابعة 398
- اختبار T لعينات تابعة لفرق المتوسط 400
- اختبار T لعينتين غير مستقلتين باستخدام SPSS 405
- تفسير مخرجات لاختبار T لعينتين ثنائيين 406
- أسئلة للمراجعة 410
- الهوامش والمصادر 413

الفصل الرابع عشر: فروق الدلالة: اختبار F لأكثر من وسطين متساويين
تحليل التباين ANOVA 415
1- اختبار الفرض لأكثر من عينتين: الفكرة العامة 415
2- تحليل التباين الأحادي: (اختبار F) 419
3- إجراء اختبار F باستخدام SPSS 426
4- تفسير مخرجات SPSS لتحليل التباين 427
- أسئلة للمراجعة 433
- هوامش والمصادر 435

الجزء الخامس: الإحصاءات الوصفية المتعددة 437

الفصل الخامس عشر: التوسع في جداول التقاطع: إضافة متغيرات التحكم 439
- مقدمة 439
1- العلاقة المباشرة 440
- إجراءات توليد جداول التقاطع مع التحكم في المتغيرات باستخدام SPSS 444
- تفسير مخرجات SPSS لجداول التقاطع مع التحكم في المتغيرات 446
2- العلاقة الكاذبة أو الدخيلة 446
- جاما الجزئية 447
- العلاقة الكاذبة أو الدخيلة 449
- العلاقة المشروطة 450
- الخلاصة 456
- أسئلة للمراجعة 457
- الهوامش والمصادر 459

الفصل السادس عشر: الارتباط الجزئي والانحدار المتعدد 461
- مقدمة 461
أولاً: الارتباط الجزئي 461

- حساب الارتباط الجزئي 462
- أنماط العلاقات 462
  • العلاقات المباشرة 462
  • العلاقات الكاذبة والدخيلة 463
  • التفاعل 464
- حساب وتفسير معامل الارتباط الجزئي 464
  • الارتباط الجزئي باستخدام SPSS 468

ثانياً: تحليل الانحدار المتعدد 469
- مراجعة الانحدار الثنائي 469
- مقدمة للانحدار المتعدد 472
- إجراءات الانحدار المتعدد باستخدام SPSS 476
- اختبار الدلالة للنموذج المتعدد 481
- الانحدار التدرجي 481
- إجراءات الانحدار المتعدد التدرجي باستخدام SPSS 484
- الافتراضات التي ينبغي مراعاتها عند استخدام الانحدار المتعدد 486
- حدود استخدام الانحدار المتعدد والارتباط 487
- أسئلة للمراجعة 489
- الهوامش والمصادر 491

الجزء السادس: الإحصاءات اللابارامترية 493

الفصل السابع عشر: اختبار مربع كاي لفروق الدلالة لعينة واحدة لتوزيع تكراري 495
- مقدمة 495
- اختبار مربع كاي لحسن المطابقة 498
- إجراء اختبار مربع كاي لحسن المطابقة باستخدام SPSS 504

- تفسير مخرجات SPSS لمربع كاي 505
اختبار مربع كاي لحسن المطابقة للحالة السوية (الاستواء) 505
- خلاصة 509
- أسئلة للمراجعة 510
- الهوامش والمصادر 512

الفصل الثامن عشر: اختبار مربع كاي للاستقلال 513

- مقدمة 513
- اختبار مربع كاي واختبارات الدلالة الأخرى 513
  - الإحصاء الوصفي المستخدم في وصف البيانات الخام 514
  - عدد العينات التي ينبغي مقارنتها 515
- الاستقلال الإحصائي 516
- حساب وتفسير إحصاء الاختبار لجداول مربع كاي للاستقلال 517
- توزيعات مربع كاي 526
- إجراء اختبار مربع كاي باستخدام برنامج SPSS 528
- تفسير مخرجات SPSS لمربع كاي 530
- الافتراضات والقيود في استخدام مربع كاي 533
- اختبار أحادي الجانب واختبار ثنائي الجانب 537
- قياس حجم التأثير لاختبار مربع كاي للاستقلال 537
- معامل فاي ϕ و CRAMERŚ V 537
- القوة 541
- تطبيقات خاصة لاختبار مربع كاي 543
- مربع كاي وارتباط بيرسون (r) 543
- مربع كاي والمقاييس المستقلة t وأنوفا 545
- معامل فاي ϕ 547
- اختبار الوسيط للعينات المستقلة 547

- تطبيقات خاصة لاختبارات مربع كاي 543
- أسئلة للمراجعة 551
- الهوامش والمصادر 555

الفصل التاسع عشر: اختبار توزيع ذي حدين لعينة واحدة 557

- مقدمة 557
- البيانات الاسمية 558
- البيانات الترتيبية والبيانات ذات المسافات والنسبية 560
- توزيع المعاينة لنسب عينة 560
- اختبار لنسبة توزيع اختبار ثنائي الحد 562
- إجراء اختبار ثنائي الحد باستخدام SPSS 565
- تفسير مخرجات SPSS لاختبارات عينة واحدة لتوزيع ثنائي الحد 566
- تقدير نسبة مجتمع 567
- الاستدلال باستخدام فترة الثقة للنسب 568
- العلاقة بين مربع كاي واختبار Z ثنائي الحد 570
- أسئلة للمراجعة 575
- الهوامش والمصادر 576

الفصل العشرون: الأساليب الإحصائية للبيانات الترتيبية 577

1- اختبار الفرض 577
أولاً: اختبار مان- وتني كبديل لاختبار مقاييس T 580
- الفرضية الصفرية المرتبطة باختبار مان-وتني U 581
- حساب قيمة اختبارات U 582
- الاختبارات الصفرية من خلال اختبار مان - وتني U 584
- التقدير التقريبي الطبيعي لاختبار مان - وتني U 585
- الافتراضات والمحاذير لاستخدام اختبار مان - وتني U 587

ثانياً: اختبار الإشارة والرتب ولكوكسن (T) 588
الفروض المتعلقة باختبار ولكوكسن T 589
- حساب وتفسير اختبار ولكوكسن (T) 590
- الدرجات المتعادلة و درجات صفر 591
- اختبار الفرض باستخدام ولكوكسن T 592
• تجاهل الفروق التي تساوي صفراً (0) 593
• تضمين الفروق التي تساوي صفراً (0) 594
- التقدير التقريبي الطبيعي لاختبار ولكوكسن (T) 595
ثالثاً: اختبار كروسكال وليز 598
- الفرضية الصفرية لاختبار كروسكال - وليز 600
- المعادلات المرتبطة باختبار كروسكال - وليز 600
رابعاً: اختبار فريدمان: كبديل لمقاييس أنوفا المتكررة 602
- مستوى البيانات المطلوبة لاختبار فريدمان 603
- الفرض الصفري لاختبار فريدمان 604
- معادلة وحساب اختبار فريدمان 604
- الإرشادات العامة لاستخدام SPSS لحساب اختبار: مان وتني U، ولكوكسن T، كروسكال - وليز وفريدمان 607
أولاً: اختبار مان وتني U 607
• الإجراء المتبع 607
• المخرجات 607
ثانياً: اختبار ولكوكسن 608
• الإجراء المتبع 608
• المخرجات 608
ثالثاً: اختبار كروسكال - وليز 609
• الإجراء المتبع 609

- • المخرجات 609
- رابعاً: اختبار فريدمان 610
- • الإجراء المتبع 610
- • المخرجات 610
- – أسئلة للمراجعة 611
- – الهوامش والمصادر 615

الملاحــق 617

- – ملحق (1) الجداول الإحصائية 619
- – ملحق (2) مسرد لبعض المفاهيم الإحصائية الواردة في متن المصنف 633
- – ملحق (3) منظم إحصائي 646
- – ملحق (4) بعض المعادلات الأساسية الواردة في الكتاب 656

المصــادر 663

- أولاً: المراجع العربية 663
- ثانياً: المراجع باللغة الإنجليزية 664

مقدمة

يهدف هذا الكتاب "الأساليب الإحصائية الوصفية والاستدلالية في تحليل البيانات" في المقام الأول إلى تَعلُمْ الأساليب الإحصائية واستخداماتها في تحليل البيانات بشكل يسير ما أمكن الأمر. ويتيح هذا الكتاب المرجعي فرصاً تساعد طلابنا على ممارسة الأساليب الإحصائية التي يتعلمونها من خلال الأمثلة والتوضيحات الواردة في ثنايا فصول هذا الكتاب. فليقرءوا هذه الفصول جيداً بدلاً من حفظ المعادلات. وقد عَمِدتُ بشكل علمي إلى عرض الأساليب الإحصائية في السياق التصوري الذي يشرح لماذا طوِّر هذا الأسلوب الإحصائي؟ ومتى يتم استخدامه؟ فإذا تمت قراءة فصول هذا الكتاب، والمادة العلمية التي تحتويها فصوله، وزيادة فهم المفاهيم الأساسية التي تتضمنها المعادلات الإحصائية، فالقارئ لاشك في أنه سيجد أن فهم هذه المعادلات، والطريقة التي تستخدم في سياقها، كل ذلك، بالضرورة، سيسهل له عملية فهم هذه الأساليب الإحصائية الوصفية والاستدلالية. حيث إنه لا غنى للباحث عن فهم واستيعاب كلا النوعين من هذه الأساليب كي يتمكن من تنظيم، ومعالجة، وتفسير بياناته بشكل وافٍ ومتكامل. فالفهم المستنير لهذه الأساليب الإحصائية الوصفية والاستدلالية يعتمد بدون شك، على

مدى قدرة الباحث على إلمامه واستيعابه لهذه الأساليب الإحصائية، ومن ثم قدرته على توظيفها بشكل علمي دقيق.

نأمل أن يساعد هذا الكتاب في فهم عملية تحليل البيانات، وأن يكون عاملاً مساعداً على تنفيذ الأسلوب الإحصائي الذي يتم اختياره بشكل ملائم وصحيح، وأن تُفَسَّر النتائجُ المُتَوَصَّلُ إليها بشكل دقيق يتلاءم والأسئلة البحثية المطروحة للبحث والاستقصاء.

من خلال هذا الكتاب حاولنا أيضاً أن نبين جملةً من المظاهر لتطوير القدرة الإحصائية لَدَى الدارسين والباحثين، لعل أهم هذه المظاهر، أولاً: العلاقة بين الإحصاء والبيانات، وتصميم البحث. فالتركيز على عملية تصميم البحث تشكل الأساس في عملية التحليل الإحصائي. بمعنى أن تصميم البحث الجيد أو الخطة الجيدة لجمع البيانات تعتبر عملية جوهرية للبحث الجيد. فالنتائج التي يتوصل إليها الباحث من خلال التحليلات الإحصائية لا تعتبر تحليلات جيدة، إلا إذا استندت على خطة جيدة توجه عملية جمع البيانات. ثانياً: التفسيرات الإحصائية. فالتفسيرات الإحصائية تُمثل تحدياً كبيراً للمشتغلين في مجال البحث الاجتماعي. فقدرة الباحث على تفسير البيانات الإحصائية يمكن تطويرها من خلال المواجهة والخبرة. ولتوفير المواجهة، فإننا قد حاولنا بشكل دقيق، من خلال الأمثلة الواردة في متن هذا الكتاب أن نبين المعنى الإحصائي فيما يتعلق بالسؤال الأساسي المطروح. ولتوفير الخبرة، فقد حاولت في نهاية كل فصل أن أضمن بعض الإجراءات المتعلقة باستخدام مجموعة Spss، وحاولتُ أن أبين الكيفية التي تتم بها عملية تفسير مخرجات هذه المجموعة الإحصائية. ثالثاً: ولكي تكون لدى الدارس الكفاءة في إجراء العمليات الحسابية الإحصائية، فقد ضمنت في نهاية كل فصل من فصول هذا الكتاب العديد من المسائل التي صممت أساساً للمراجعة حتى يكون الدارس على دراية بما تم طرحه من خلال هذه الفصول. رابعاً: ومن المظاهر الأخرى لتطوير القدرة الإحصائية لدى الباحثين والدارسين، فقد حاولت في سياق هذا الكتاب أن أطور من قدرة الباحثين والدارسين على قراءة التقارير والأدبيات العلمية والمهنية في مجال العلم الاجتماعي. فالشخص الواعي بالأساليب الإحصائية يمكنه استيعاب ونقد، وتقدير التقارير العلمية التي أعدت من قِبل الآخرين. هذه هي تلك المظاهر التي تشكل الإطار العام لهذا الكتاب. وفي المحصلة النهائية، فإن الهدف الأساسي من

تصميم هذا الكتاب المرجعي، هو مساعدة طلابنا، والمهتمين بمجال البحث العلمي بشكل عام، والبحث في مجال العلم الاجتماعي بشكل خاص، في فهم الأساليب الإحصائية الوصفية والاستدلالية واستخداماتها الملائمة في تحليل البيانات، ومنحهم الثقة اللازمة للتعامل مع هذه الأساليب الإحصائية واستخداماتها بشكل علمي دقيق.

محتويات الكتاب:

يتألف الكتاب من عشرين فصلاً تغطي بدءاً من الإحصاءات الوصفية وحتى الإحصاءات الاستدلالية المتقدمة. وينقسم الكتاب في ذلك إلى ستة أجزاء. يعرض الجزء الأول الإحصاءات الوصفية الأحادية: الإحصاء والبيانات، وتصميم البحث، المتغيرات والقياس، تحليل البيانات التكرارية، مقاييس النـزعة المركزية ومقاييس التشتت، المنحنى الطبيعي. أما الجزء الثاني فيتناول الإحصاءات الوصفية الثنائية: تحليل جداول التقاطع – استقصاء العلاقة بين متغيرين (الوصف الجدولي لبيانات مقاسة على المستويين الاسمي والترتيبي) – الوصف العددي للبيانات الاسمية، مقاييس التطابق الثنائية، والتطابق بين متغيرات تم قياسها على المستوى ذي المسافات والنسبي. ويتناول الجزء الثالث الإحصاءات الاستدلالية: تحليل البيانات المستخدمة من عينة واحدة، توزيعات المعاينة، التقدير، فترات الثقة، اختبار الفروض: اختبار Z لمتوسط عينة واحدة، اختبار t لمتوسط حسابي لعينة واحدة. في حين يركز الجزء الرابع على الإحصاءات الاستدلالية لعينتين أو أكثر المتمثل في: اختبار t لعينتين ذات وسطين حسابيين متساويين، فروق الدلالة: اختبار F لأكثر من وسطين، ويلقي الجزء الخامس الضوء على الإحصاءات الوصفية المتعددة من خلال التوسع في جداول التقاطع بإضافة متغيرات التحكم، الارتباط الجزئي والانحدار المتعدد. أما الجزء السادس والأخير، فيناقش الإحصاءات الاستدلالية المتمثلة في اختبار مربع كاي لفروق الدلالة لعينة واحدة لتوزيع تكراري، اختبار مربع كاي للاستقلال، اختبار التوزيع ذي الحدين لعينة واحدة، والأساليب الإحصائية للبيانات الترتيبية كاختبار مان – وتني، واختبار ولكوكسن، وكروسكال – وليز، وفريدمان.

وبعد، فإنني أضع هذا الكتاب بين أيدي الباحثين والدارسين وعامة المهتمين بالشأن

الاجتماعي، راجياً أن يلمسوا فيه نوعاً من الفائدة العلمية، وأسألهم ألا يضنّوا عليّ بأية ملاحظات قد يقترحونها أو تصويب قد يرونه فالكمال لله وحده.

وختاماً، فإنني أتقدم ببالغ الشكر لاسيما إلى كل من الأستاذ/ جمال محمود لمراجعته لمسوّدة الكتاب لغوياً والأستاذة / أسماء الشريف على عنايتها بمراجعة المسودات الأولى لهذا الكتاب، والشكر موصول إلى الآنسة / هنية العلواني على تحملها طباعة وتصحيح التجارب الطباعية في مراحلها المختلفة.

والله أدعوا أن يوفقنا جميعاً لما فيه الخير، وأن ينفع بعلمنا هذا أمتنا العربية، إنه سميع مُجيب.

د. عبد الله عامر الهمالي

جامعة بنغازي - 2011 م

الجزء الأول

الإحصاءات الوصفية الأحادية

- الفصل الأول: الإحصاء والبيانات وتصميم البحث
- الفصل الثاني: المتغيرات والقياس
- الفصل الثالث: تحليل البيانات التكرارية
- الفصل الرابع: مقاييس النزعة المركزية ومقاييس التشتت
- الفصل الخامس: المنحنى الطبيعي

الفصل الأول

الإحصاء والبيانات وتصميم البحث

مقدمة

يلعب التحليل الإحصائي الإحصائي دوراً كبيراً في العلم الاجتماعي والسلوكي، فقد لا تخلو أية مجلة علمية أو كتاب في هذين الحقلين إلاّ والاستخدامات الإحصائية حاضرة سواء كانت هذه الإحصاءات متعلقة باختبارات الدلالة أو الارتباطات أو الإحصاءات الوصفية. إن هذا الاستخدام ليس مستبعداً، فالإحصاء كأداة يستخدمها علماء العلم الاجتماعي والسلوكي هي نفس تلك الاستخدامات لدى علماء العلم الطبيعي. وباستخدام العلماء للأدوات الإحصائية فإنهم بذلك قد وصلوا إلى تحقيق أهداف عالية من حيث قدرتهم على بناء نظريات علمية قادرة على تفسير الفعل الإنساني، أو إيجاد أرضية لبناء حالة موضوعية تساعد على إيجاد برامج لخدمات اجتماعية مستمرة.

لاشك أنه بدون الإحصاء، قد نجد أنفسنا سُبَّحاً في بحر من الوقائع والأرقام، فالعقل الإنساني عادة قد لا يمكنه استيعاب هذه الحقائق والأرقام بالكيفية التي هي عليها. يجدر بنا القول بأن الباحثين في مجال العلوم الاجتماعية والسلوكية عليهم – وبشكل روتيني –

التعامل مع كم هائل من المعلومات (البيانات)، ومن هنا جاء دور الإحصاء لتلخيص وتنظيم وتحليل البيانات الرقمية بشكل مرتب وواضح يسهل فهمه بكل سهولة ويسر.

إن الاهتمام بالمعرفة الإحصائية، أصبح اليوم ضرورة ملحة، فالباحث الاجتماعي ينبغي عليه الإلمام بشكل كبير بالبرامج الإحصائية التي أصبحت اليوم مُيَسَّرة، فالباحث الُمزَوَّد بالرُّزَم الإحصائية دون فهم عميق بالكيفية التي يعمل بها الإحصاء، قد تكون خطيرة حقاً. فالبرامج الإحصائية لا تعطي الباحث أكثر مما يدخله من بيانات، ومن هنا يأتي دور الباحث المعرفي لتقييم ما تعنيه هذه المخرجات، أو بشكل أكثر دقة، ما هو المغزى الكامن وراء هذه المخرجات الإحصائية؟ فإجراء العمليات الإحصائية ليس إلا مجرد عمليات إحصائية ولكن الأمر يكمن في ما مغزى هذه العمليات الإحصائية؟ أو ما هي الدلالة وراء هذه العمليات الإحصائية؟

لاشك في أن الحاسبات قد تطورت وأن الباحث بإمكانه الحصول على التقنيات الإحصائية الأكثر تعقيداً في لمح البصر بدلاً من الأيام أو الأسابيع. وبالرغم من أن الحاسبات قد قللت من الحاجة إلى فهم آلية التحليل الإحصائي، فإنها تظل غير قادرة على اختيار أنسب أنواع الإجراء الإحصائي لمجموعة من الحالات المحددة، ولن تكون هذه الحاسبات قادرة على الإجابة على الأسئلة التي يطرحها البحث. فالإجابات تكون لدى الباحث أكثر ما تكون قيماً تم طبعه (المخرجات).

باختصار، إنَّ الاستخدام الفعال للإحصاء يحتاج إلى فهم تصوري للتحليل الإحصائي[(1)]

إن الهدف من هذا الكتاب هو التركيز على الاختبارات الإحصائية واستخداماتها في مجال البحوث الاجتماعية والسلوكية. فقد خصصت فصول هذا الكتاب إلى فرد فصول مستقلة لكل اختبار من الاختبارات الإحصائية حتى لا يقع الباحث في أي إرباك أو تشويش في تطبيقه لهذا الاختبار أو ذاك. كما صُمِّم هذا الكتاب أيضاً كدليل شامل لاستخدام هذه الاختبارات الإحصائية في العلوم الاجتماعية والسلوكية.

لقد قُسِّم هذا الكتاب إلى مجموعتين من المهارات الإحصائية. أولهما: مهارات حساب وتفسير الإحصاءات الوصفية، وثانيهما: المهارات المتعلقة بحساب وتفسير الإحصاءات

الاستدلالية، فالإحصاءات الوصفية هي تلك الأعداد التي تلخص مجموعة من البيانات مع وصف أنماط في تلك البيانات. وببساطة فإن بعض الإحصاءات تصف الصفات الغالبة كالنوع، والعمر... الخ، في حين إحصاءات أخرى تصف العلاقات خاصيات النوع، الاتجاه، النوع والدخل، إن الإحصاءات الوصفية لا تقودنا إلى معرفة ما إذا كانت العينات حقيقة تمثل المجتمع الذي أخذت منه هذه العينات.

على الجانب الآخر، إن الإحصاءات الاستدلالية تساعد في تقييم ما إذا كانت التعميمات أو الاستدلالات من العينات إلى المجتمع ملائمة أم لا؟ ولفهم الإحصاءات الوصفية والإحصاءات الاستدلالية دعنا نبدأ باستعراض عملية إجراء البحث وذلك من خلال مراجعة بعض المفاهيم الأساسية لتصميم البحث الاجتماعي.

إن عملية تصميم البحث تشكل أساس التحليل الإحصائي حيث لا توجد أي كمية من الاستخدامات يمكن أن تعوض ما إذا كان هناك قصور منطقي في تلك البيانات التي تم جمعها للتحليل. بمعنى آخر، أن تصميم البحث الجيد أو الخطة لجمع البيانات تعتبر عملية جوهرية للبحث الجيد.

إن النتائج التي نتوصل إليها من التحليلات الإحصائية لا تعتبر تحليلات جيدة إلا إذا استندت على خطة توجه عملية جمع البيانات. ومن هنا رأينا أنه من الضرورة بمكان اعتماد فصل يتعلق بتصميم البحث كجزء من فصول هذا الكتاب.

إن مَعْرِفتَنَا بالإحصاء هي أيضاً مَعْرِفَتُنَا بتصميم البحث. فالباحث الذي يعتمد بشكل كبير على اختبارات الدلالة (مثل اختبار t، واختبارات التباين... الخ) سوف يميل إلى النظر إلى المشكلات البحثية في إطار المتغيرات المستقلة والمتغيرات التابعة، المجموعات التجريبية والمجموعات الضابطة؛ ومن هنا يمكننا القول بأنه كلما كان ذا معرفة بالتقنيات الإحصائية كان أكثر مرونة في تصميم البحث. ينبغي على الباحث أن يكون قادراً على توظيف الإحصاء بشكل فعّال لتنظيم بياناته وتقيمّها وتحليلها.

إنه بدون فهم جيد لأسس التحليل الإحصائي، فالباحث سوف يكون غير قادر لبيان ما إذا كانت البيانات منطقية. وعليه يمكننا القول بأنه إذا لم تطبق التقنيات

الإحصائية بشكل ملائم، فإن البيانات المجمعة ستبقى غير ذات جدوى، ومن هنا تأتي أهمية الإحصاء الذي لا غنى عنه في العلوم الاجتماعية، فالإحصاء يزودنا بتقنيات جد مهمة في تقييم الفروض واختبار النظرية. كذلك يساعد الإحصاء علماء العلم الاجتماعي في قدرتهم على إجراء البحوث الكمية. فالبحث الكمي يقوم أساساً على تحليل المعلومات العددية أو البيانات. فالباحثون يستخدمون التقنيات الإحصائية لتنظيم ومعالجة البيانات ليتمكنوا من اختبار الفروض وتطوير النظريات أو صقلها وكذلك فهمنا للعالم الاجتماعي من أجل إصلاحه. ومن هنا يمكننا القول، بأن الباحث الذي يكون لديه خلل في تصميم بحثه، فإنه بالتالي لا يستطيع توظيف الإحصاء بشكل فعّال. فالإحصاء لا يكون بديلاً للتصورات الدقيقة والخطة المحكمة أو الاستخدام الخلاق للنظرية. فالإحصاء ليس منقذاً لتصميم بحث رديء.

عملية البحث الاجتماعي:

تمر عملية البحث الاجتماعي في العلم الاجتماعي بالمراحل التالية:

تحديد أهداف البحث:

إن أول خطوة تمر بها عملية البحث الاجتماعي أن يحدد الباحث ماذا يريد، أَيْ ما الذي يريد دراسته مبرزاً السؤال البحثي الذي سيقوده إلى الاستقصاء والبحث.

مراجعة الأدبيات المتعلقة بالمشكلة المطروحة:

ففي هذه المرحلة يُتَطَلَّبُ من الباحث أن يضع المشكلة المطروحة في الإطار العام المرتبط بالنظرية والبحث، مراجعة الدراسات السابقة حول الموضوع المطروح، هل الدراسات السابقة تتباين فيما بينها؟ هل المعلومات المتوفرة حول الموضوع يمكن للباحث أن يعالجها بشكل جيد في إطار دراسته الحالية؟

صياغة الفروض:

إن مراجعة الأدبيات والدراسات المتعلقة بالموضوع المطروح تساعد في صياغة فروض الدراسة.

القياس:

تشمل هذه الخطوة المتغيرات الأساسية للدراسة. وكيف يمكن تعريف هذه المتغيرات وقياسها؟ هل التعريفات والقياسات الواردة في هذه الدراسة تختلف عن تلك التعريفات والقياسات الواردة عند الآخرين، أم أنها تصب في نفس السياق؟ فإذا ما تم فعلياً تطوير أداة القياس (الاستبيان على سبيل المثال) أم أن الباحث سيستخدم أداة قياسية تم تطويرها من قِبَل الآخرين، فإنه من المناسب أن يضمّن هذا القياس الذي تم استخدامه من قِبل الباحث في ملاحق دراسته.

طرق جمع البيانات:

هنا ينبغي على الباحث أن يحدد ما هي الطرق المناسبة لجمع البيانات المتعلقة بالدراسة؟ هل يسعى الباحث إلى إجراء دراسة تجريبية أم مسح أم ملاحظة مباشرة أم أن الباحث يسعى للقيام بدراسة ميدانية أم أنه يريد التركيز على تحليل بيانات إحصائية تم جمعها من قِبل باحثين آخرين. - تجدر الإشارة إلى أن الباحث يمكنه استخدام أكثر من طريقة لجمع البيانات -.

تحليل البيانات:

في هذه المرحلة ينبغي على الباحث أن يشير إلى نوعية التحليل الذي سيقوم به إشارة واضحة إلى غرض ومنطق هذا التحليل. هل الباحث يرغب في تحليل بياناته بحيث لا تتجاوز الوصف؟ أم أن الباحث لديه الرغبة في تفسير لماذا تحدث الأشياء بهذه الكيفية؟ فعلى سبيل المثال، لماذا يسجل بعض الطلاب بعض الاتجاهات الإيجابية حول قضية ما دون غيرهم؟ ما هي المتغيرات التفسيرية التي يتوجب على الباحث وضعها في الاعتبار عند عملية تحليل البيانات؟ وكيف يمكن للباحث أن يعرف ما إذا كانت لديه تباينات تفسيرية دقيقة؟

الدعوة إلى التدقيق:

وذلك من خلال إتاحة الفرصة للآخرين للتعرف على البحث والتحليل. إضافة إلى

استعداد الباحث العلمي لمراجعة إجراءات عملية البحث التي تم اتباعها، البيانات التي تم جمعها والنتائج التي توصل إليها البحث.

الأسس المتعلقة بتصميم البحث:

المجتمع، عناصر المجتمع، ووحدات التحليل:

ريثما يحدد الباحث الأسئلة البحثية، عندئذ يتوجب عليه أن يقرر من هم الأشخاص المستهدفون بالدراسة للإجابة على الأسئلة التي يطرحها البحث. فالباحث عليه أن يقوم بجمع البيانات حول خصائص هؤلاء الأفراد أو المجموعات أو المؤسسات أو أي وحدة تحليل أخرى.

المجتمع:

المجتمع هو مجموعة من العناصر التي يسعى الباحث للحصول على بيانات منها أو حولها. فالعنصر هو وحدة كينونية في المجتمع. فالوحدات المفردة تؤلف مجموعة يمكن أن تكون أفراداً – مجموعات – منظمات أو مؤسسات أو أي وحدة اجتماعية منظمة (مثل القرية، المدينة) رسمية أو غير رسمية.

إن بعض المجتمعات تحتوي على مجموعة من الأفراد تمثل عنصراً لحالة منفردة من المجتمع وهي الفرد. فالمجتمع كما أشرنا، يشير إلى مجموعة من الأفراد في حين أن مجتمعات أخرى يمكن أن تحتوي على مجموعة دول، أمم، قرى، إلى آخره، والتي تتحدد بعنصر مفرد يمثل المدينة أو القرية. فالباحث عندما يرغب في دراسة معدلات الجريمة في المدن الكبرى في ليبيا، على سبيل المثال، فإن بإمكانه دراسة المجتمع (مجموعة) مثل المدن الكبرى: طرابلس، بنغازي، سبها، مصراته. أيضاً قد يرغب الباحث في دراسة معدلات الفقر في العالم، فإنه بذلك يستطيع دراسة المجتمع العالمي وهنا قد تكون وحدة التحليل عنصراً مفرداً مثل المكسيك أو أي دولة أخرى في العالم، إذاً الباحث يقوم بجمع البيانات من العناصر المكونة للمجتمع للكشف عن الأنماط في واحد أو أكثر من خصائص المجتمع ككل. إن السؤال

الذي يُطْرَحُ هنا هو: ما هي خصائص ذلك المجتمع الذي يرغب الباحث في معرفتها؟ إِنَّ الإجابة على مثل هذا السؤال تقودنا إلى الحديث عن وحدة التحليل.

وحدة التحليل:

إن وحدة التحليل هي كينونة محددة سعى الباحث لمعرفة شيء ما حولها. تجدر الإشارة إلى أنه عادة ينظر إلى عناصر المجتمع ووحدات التحليل كشيء واحد، ففي معظم الأسئلة التي تطرح في المسوح الاجتماعية الكبرى فإن الفرد يمثل وحدة التحليل في هذه المسموح. فالبيانات عادة ما تُجْمَعُ من خلال المقابلة الشخصية لهؤلاء الأفراد وجهاً لوجه لمعرفة الخصائص العامة لهؤلاء الأفراد مثل: النوع، الوضع الاجتماعي، المستوى التعليمي، الدخل... الخ، فعلى سبيل المثال، قد توجد مجموعة أخرى من الأسئلة تتعلق بالأسرة، مثل: الدخل الأسري، وعدد الأطفال في الأسرة ففي مثل هذه المتغيرات فإن وحدة التحليل هي الأسرة.

العينات وإطار المعاينة:

نادراً ما يلجأ الباحث إلى دراسة المجتمع ككل وذلك لكبر حجمه، ومن هنا يلجأ الباحث إلى اختيار عينة من ذلك المجتمع يتم اختيارها بطريقة عشوائية بحيث تمثل هذه العينة الخصائص العامة للمجتمع المدروس التي سحبت منه هذه العينة. ولكي تكون العينة ممثلة للمجتمع ينبغي على الباحث سحب العينات طبقاً للقواعد الاحتمالية. فالعينات الاحتمالية هي تلك العينات التي يتم سحبها بطريقة تتيح فرصاً متساوية أمام جميع وحدات المجتمع المدروس. وقبل سحب العينة ينبغي على الباحث أن يكون لديه إطار للمعاينة أو قائمة تحتوي على كل العناصر المكونة للمجتمع. ويتخذ إطارُ المعاينة أشكالاً مختلفةً، كالمجموعة الإحصائية والتعداد السكانيِّ، ودليل المدينة، وقوائمَ بأسماءِ الوحداتِ المطلوبِ دراستُها. وعندما يتوفر إطار المعاينة لدى الباحث عندئذ يكون بإمكانه سحب عينة عشوائية ومن التقنيات الشائعة الاستخدام لسحب عينات احتمالية، المعاينة العشوائية البسيطة، والمعاينة المنتظمة.

المعاينة العشوائية البسيطة:

وتعني المعاينة العشوائية البسيطة المعاينة التي تتيح فرصاً متساوية أمام جميع وحدات الظاهرة المدروسة. والمعاينة العشوائية بأبسط إجراءاتها يمكن من خلالها أن يكون لدى الباحث قائمةٌ حيث يتم تسجيل كل العناصر في أوراق منفصلة ثم تخلط هذه الأوراق ويختار من بينها العدد المطلوب. ويمكن للباحث أن يستخدم نفس الإجراء إلكترونياً من خلال برامج الحاسب الآلي لتوليد عينات عشوائية. أو يمكنه اللجوء إلى نفس الإجراء من خلال الجداول العشوائية بأن يختار الباحث رقماً عشوائياً من أي مكان في الجدول.

المعاينة المنتظمة ببداية عشوائية:

إن المعاينة المنتظمة ببداية عشوائية تشبه المعاينة العشوائية البسيطة، عدا أن المعاينة المنتظمة ببداية عشوائية أن فيها العنصر الأساسي يتم اختياره باستخدام جدول الأرقام العشوائية، وعندما يحدد الباحث المسافة يبدأ في الاختيار العشوائي.

عملية المعاينة متعددة المراحل:

تجدر الإشارة هنا إلى أن هناك إجراءاتٍ معاينة أكثر تعقيداً من حيث التصميم، ذلك أن هذا النوع من المعاينة يمر بعدة مراحل. فالباحث بداية عليه أن يقسم الإطارات التي من خلالها يود سحب عيناته. فالتقسيم إلى فئات يتطلب تقسيم العناصر وفقاً لخصائص معينة أو معيار معين مثل: النوع، أو الخلفية الحضرية أو الأثنية. ويتطلب من الباحث أن يقسم إطارات المعاينة ليضمن أن العينات التي يختارها ستكون ممثلة لواحد أو أكثر من الخصائص المهمة في المجتمع الذي يود دراسته. على سبيل المثال، يمكن للباحث أن يقسم الإطار إلى فئات طبقاً للنوع مولداً قوائم منفصلة للذكور والإناث، وبعد ذلك، يمكنه سحب عينة مستخدماً إما العينة العشوائية أو العينة المنتظمة من كل قائمة. إن تقسيم المجتمع إلى ذكور وإناث يسمح للباحث بأن العينة المختارة ستتضمن الذكور والإناث، نسبة لتمثيل الذكور والإناث في المجتمع ككل.

ومن استراتيجيات المعاينة الأخرى، عينة التجمعات. ويستخدم هذا النوع من المعاينة

عندما لا يكون في مقدور الباحث الحصول على إطارٍ جيدٍ وقائمةٍ للعناصر المكونة للمجتمع، وكنتيجة لذلك، ينبغي عليه في هذه الحالة أن يولد قائمة بالتجمعات أو كينونات اجتماعية منظمة تُشْبِع فضول الباحث. على سبيل المثال، إن الباحث الذي يرغب في إجراء مسح عما يريده الشباب في المجتمع الليبي، وأنه ليس بإمكانه الحصول على قائمة لكل الشباب الليبي بليبيا، فإنه في هذه الحالة ينبغي عليه أن يجد إطاراً لكل الوحدات الجغرافية التي يمكن من خلالها الوصول إلى الشباب الليبي، سواء أكانت هذه الوحدات الجغرافية مدن أو قرى أو نجوع، ومن خلال هذا الإطار يمكنه عندئذ سحب عينة عشوائية ومن هذه التجمعات مستخدماً المعاينة العشوائية أو المعاينة المنتظمة. وبعد ذلك يمكن للباحث أن يختار عشوائياً الأفراد داخل كل مجتمع للإجابة على الأسئلة التي يطرحها الباحث.

المتغيرات:

تلعب المتغيرات دوراً أساسياً في عملية البحث الاجتماعي ومن هنا ينبغي على الباحث فهم هذه المفاهيم التي تمثل الخصائص التي سَتُجْمَعُ البيانات منها. فالمتغيرات عبارة عن أي مظهر لوحدة التحليل التي يمكن أن تتباين من وحدة تحليل إلى أخرى. فإذا كانت وحدة التحليل الفرد، فإن المتغيرات يمكن أن تحتوي على خصائص مثل: العمر، النوع، عدد سنوات الدراسة، الديانة،... الخ من الخصائص. فالباحث عندما يجري بحثه فإنه في العادة يرغب في التعامل مع عدد كبير من المتغيرات فهو يقوم بجمع عدد لا بأس به من البيانات الديموغرافية والخلفية الاجتماعية للمبحوث الخاضع للدراسة.

إضافةً إلى هذه المتغيرات التي أشرنا إليها، فإن الباحث يقوم بجمع بيانات حول المتغيرات المتعلقة بالوضع الاجتماعي، والوضع المهني،... الخ، كما يمكنه أيضاً جمع بيانات حول اتجاهات الأفراد نحو قضايا معاصرة مثل الاتجاه نحو تولي المرأة مناصب إدارية عليا، أو الاتجاه نحو توزيع الثروة على الليبيين.

الفئات، والقيم، والبيانات:

يقوم الباحث بجمع بياناته عن كل متغير داخل الفئات فالسؤال حول النوع أو السؤال

حول العمر مثلاً، فإذا ما أجاب المبحوث بأنه ذكر فهنا نطلق على هذه الإجابة فئة متغير النوع أو فئة متغير العمر إذا ما حدد المبحوث عمره بالسنوات. إذاً المتغير يحتوي على الأقل فئتين (إذا تضمن أي متغير فئة واحدة فإن ذلك لن يؤدي إلى أي تباين بين المبحوثين) وعندما لا تتباين الخصائص من مبحوث لآخر، فإننا نطلق على ذلك "متغير ثابت" Constant Variable. فالباحث في بعض الأحيان يخصص أعداداً لفئات المتغيرات. يمكن للباحث أن يخصص رقم (1) لكل الإجابات التي أجابت (ذكور) عند طرح سؤال النوع. وكل واحد أجاب على سؤال النوع (أنثى) يخصص له رقم (2). وهذه الأرقام التي تم تخصيصها من قِبل الباحث لفئات المتغير يطلق عليها القيم Value.

إن البيانات المتعلقة بكل متغير تحتوي على إجابات محددة تمّ الحصول عليها من خلال السؤال البحثي المطروح. فعلى سبيل المثال، إن البيانات المتعلقة بالنوع هي إجابات محددة تم الحصول عليها عند طرح السؤال المتعلق بالنوع. والبيانات حول العمر هي تلك الإجابات المحددة التي تم الحصول عليها عند طرح السؤال حول العمر [(2)].

الفروض:

في بعض الأحيان يقوم الباحث بجمع البيانات حول متغيرات لغرض بسيط وهو وصف إجابات المبحوثين (خصائص العينة) كنسبة الذكور إلى الإناث أو نسبة المبحوثين المتزوجين أو نسبة المتعلمين في مقابل غير المتعلمين. لاشك أن الباحث قد لا يقف عند هذا الحد بوصفه لخصائص العينة، وإنما قد يرغب في معرفة العلاقة بين المتغيرات هل الذكور في المتوسط أكثر تعليماً من الإناث؟ هل صغار السن هم أكثر استعداداً لتقبل دور المرأة الجديد من الكبار؟... الخ.

تجدر الإشارة إلى أن الباحث عادة ما يكون في ذهنه بعض التخمينات حول العلاقة بين المتغيرات، وأن هذه التخمينات قد استمدها الباحث من خلال المعرفة النظرية، أو من خلال الدراسات التي أجريت من قبل أو يمكنه استخلاص هذه التخمينات من خلال التجارب الشخصية أو من خلال المشاهدات المباشرة كوّنها الباحث حول ظاهرة ما. وعندما يعبر عن هذه التخمينات أو العلاقات في عرض يصف العلاقة على الأقل بين

متغيرين، يعني هذا أن الباحث قد طوّر فرضية حول هذه العلاقة. ولنأخذ الفرض التالي كمثال على ذلك:

- الإناث أكثر ميلاً من الذكور نحو تنظيم الأسرة.

إن هذا الفرض يبين العلاقة بين النوع والاتجاه نحو تنظيم الأسرة، وبالتالي نجد أن هذه الفرضية فرضية محددة حول طبيعة هذه العلاقة. أي كيف ترتبط المتغيرات بعضها مع البعض الآخر (الإناث أكثر ميلاً من الذكور نحو تنظيم الأسرة). ولكي نختبر صحة هذه الفرضية، ينبغي علينا جمع البيانات من مجموعة الذكور ومجموعة الإناث من خلال طرح السؤال المتعلق بالاتجاه نحو تنظيم الأسرة.

أضف إلى ذلك يمكن للباحث أن يذهب بالفرض أبعد من ذلك من خلال بيان العلاقة التي تشمل متغيراً ثالثاً. فعلى سبيل المثال، يمكننا افتراض أن الرجال أكثر حرية في التعامل مع المحيط الاجتماعي الخارجي من النساء، لاسيما وأن الرجال لديهم فرصة الخروج خارج المنزل أكثر من النساء. وإذا ما صدقت هذه الفرضية، فإننا نتوقع أن نجد نسبة الرجال أكثر من نسبة النساء حيث لديهم الحرية في التعامل مع المحيط الخارجي.

المتغيرات المستقلة والمتغيرات التابعة:

من خلال العلاقة الفرضية المحددة نستطيع الحديث عن الكيفية التي ترتبط بها المتغيرات – أي أن المتغير الذي يُفْتَرَض أن له تأثيراً على المتغير الآخر – فالمتغير الذي يحدث التأثير يطلق عليه "المتغير المستقل". في حين أن المتغير الذي يتأثر يُطْلَق عليه "المتغير التابع". بكلام آخر، إن المتغير التابع هو المتغير الناتج عن تأثير المتغير المستقل. إذاً المتغير المستقل هو المتغير التفسيري الذي يُفْتَرَض أنه يؤدي إلى التباين في قيم المتغير التابع؛ وبالتالي يكون المتغير التابع النتيجة المتوقعة للمتغير التفسيري. على سبيل المثال، إذا رغب الباحث في دراسة العلاقة بين التعليم والاتجاه نحو مشاركة المرأة في الحياة العامة، فإنه قد يصل إلى نتيجة مفادها: أن الأفراد ذوي المستويات التعليمية العالية يسجلون درجات إيجابية نحو مشاركة المرأة في الحياة العامة. ويمكن توضيح هذه العلاقة بالشكل التالي:

التعليم ←———— اتجاهات إيجابية نحو مشاركة المرأة في الحياة العامة

(متغير مستقل) (متغير تابع)

مما سبق، لاحظ أننا لم نستخدم كلمة "سبب" لعرض هذه العلاقة. ولما كان متغير التعليم اُفتُرِضَ أن له تأثيراً على اتجاهات الأفراد نحو مشاركة المرأة في الحياة العامة، فإننا بذلك لا نستطيع أن نفترض أن تعليم شخص ما يسبب أن يسجل هذا الفرد درجات إيجابية نحو مشاركة المرأة في الحياة العامة، بالرغم من أننا في بعض الأحيان نعتقد أن العلاقة بين المتغير المستقل والمتغير التابع هي علاقة سببية أي علاقة سبب ونتيجة. وعندما يتبين لنا أن أحد الخصائص المرتبطة بأحد المتغيرات (تعليم عالٍ) عندئذٍ نميل إلى إيـجاد خاصية محددة للمتغير الثاني كذلك (مواقف إيجابية حول مشاركة المرأة في الحياة العامة). أضف إلى ذلك، أنه في بعض الأحيان يصعب علينا تحديد أي من المتغيرين يمكن اعتباره مستقلاً وأي منهما نعتبره تابعاً.

يتضح في بعض المواقف أنه لا هذا يمكن نعته بأنه متغير مستقل ولا ذاك متغير تابع عند الحديث عن العلاقة بينهما. والسؤال الذي يمكن طرحه في مثل هذه المواقف هو: ماذا نعمل؟

للإجابة على هذا السؤال يمكننا أن نسوق مجموعة أدلة عامة يمكن استخدامها في كيفية التعامل مع المتغيرات [3]:

منطق العلاقة بين المتغيرات:

ينبغي على الباحث أن يعي أن هناك سبباً منطقياً في تفكيره بأن المتغير المستقل يمكنه التأثير في المتغير التابع. هل يستطيع الباحث أن يفكر في كم التأثير الذي يحدثه أحد المتغيرات على المتغير الآخر؟ على سبيل المثال، إذا كان الباحث يرغب في معرفة العلاقة بين النوع والاتجاه نحو تنظيم الأسرة، هل بإمكانه أن يفكر في الأسباب لماذا يؤثر جنس الشخص في اتجاهه نحو تنظيم الأسرة وليس بالعكس. هل الاتجاه يؤدي إلى نتيجة الاختلاف في النوع. إن الإجابة، بطبيعة الحال، لا. إذاً، من هنا يمكننا التعامل مع النوع كمتغير مستقل، والاتجاه نحو تنظيم الأسرة كمتغير تابع [4].

الوقت:

المتغير المستقل ينبغي أن يظهر في وقت سابق على المتغير التابع. ففي المثال السابق: جنس الشخص سابق على الاتجاه نحو تنظيم الأسرة باعتبار أن الهوية النوعية لأي شخص تتطور عبر الزمن.

الخصائص النوعية في مقابل الإنجاز أو الأداء:

تعتبر الخصائص النوعية دائماً متغيرات مستقلة (الخصائص النوعية هي تلك الخصائص التي يورثها شخص ما، أو هي تلك الخصائص التي لا يمتلك الشخص القدرة على التحكم فيها). ومن أمثلة الخصائص النوعية: النوع، العرق، العمر. على سبيل المثال، قد يعامل الباحث بعض المتغيرات كمتغيرات نوعية Ascribed أو سمات موروثة Inherited Traits، بالرغم من أن هذه المتغيرات قد تتغير كالوضع الاقتصادي والاجتماعي للعائلة الأصلية (العائلة التي يولد فيها الفرد).

أما فيما يتعلق بخصائص الإنجاز والأداء، كثيراً ما يتعامل معها (ليس دائماً) كمتغيرات تابعة. وتعرف الخصائص المتعلقة بالإنجاز أو الأداء بأنها تلك الصفات attributes التي يطوّرها الفرد أو يكتسبها عند الكبر. وتكتسب هذه الخصائص من خلال المركبات: كالاختيار، والجهد والقدرة أو بفعل الوصول إلى غرض أو هدف معين، فمواقف الشخص، أو التصرفات التي يتعاطى معها، هي خصائص تم إنجازها كسنوات التعليم التي أكملها، والوضع الاجتماعي والاقتصادي الذي حققه الفرد كشخص بالغ الرشد.

ثمة مواقف يكون فيها المتغيران - في علاقتهما - مستقلَّيْن عن بعضهما البعض. ففي بعض الأحيان، نجد متغيرين، لا هذا المتغير متغير مستقل ولا ذاك المتغير متغير تابع في علاقتهما ببعض، بالرغم من أنهما ربما مرتبطان ببعضهما البعض. فعلى سبيل المثال، قد يرغب الباحثون في معرفة مواقف الأفراد الذين يقومون بدراستهم، ومع ذلك، فإن مواقف أحد الأفراد ليس بالضرورة أن تسبب أو تحدث تأثيراً على موقف شخص آخر.

إن مثل هذه العلاقات يطلق عليها علاقات متماثلة Symmetrical relationship

or associations. (سيتضح هذا المفهوم بشكل جلّي عند الحديث عن مقاييس التطابق الإحصائي في فصول هذا الكتاب). في مثل هذه الحالات لا يهم الباحثَ أيٌّ من المتغيرين يمكن التعامل معه كمتغير مستقل - باعتبار أن الآخر يكون متغيراً تابعاً - في جداول التقاطع. فالباحث في كل هذه الأحوال يمكنه أن يقرر أن أحد هذين المتغيرين متغير مستقل في علاقته بالمتغير الآخر وذلك استناداً على السؤال البحثي المطروح والمحدد الذي يحاول الباحث الإجابة عليه [5].

متغيرات التحكم:

متغيرات التحكم هي تلك المتغيرات التي يمكن أن يكون لها تأثيرُ على المتغير المستقل والمتغير التابع. فإذا افترضنا أن القدرة التحصيلية للدخل مرتبطة بالوضع العائلي فهو افتراض غير صحيح، لأن عملية الدخل عملية مرتبطة بمتغير التعليم للآباء والأبناء على حد سواء [6].

أضف إلى ذلك، لو افترضت أن النساء أكثر إيجابية نحو تنظيم الأسرة من الرجال. فإذا ما أدخلنا متغير التحكم الخلفية الحضرية والريفية، وانطلقنا من أن الخلفية الحضرية / الريفية تؤثر على العلاقة بين النوع والاتجاه نحو تنظيم الأسرة. فهذا يعني أن تأثير النوع على الاتجاه قد يختلف عند الحضريين عنه لدى الريفيين.

الخلاصة:

خلال هذا الفصل بينّا خطوات عملية البحث بَدْءًا بتحديد المشكلة، مراجعة الأدبيات المتعلقة بمشكلة البحث، صياغة الفروض، القياس، تحليل البيانات، كما تم تقديم الأسس المتعلقة بتصميم البحث وتحليل البيانات الكمية: المجتمع، سحب العينة من المجتمع، عناصر ووحدات المجتمع، والعينات، والمتغيرات، وفئات المتغيرات وقيم المتغيرات، والفروض، والمتغيرات المستقلة والمتغيرات التابعة ومتغيرات التحكم.

أسئلة للمراجعة:

1- أعط مثالاً تبين فيه المتغير المستقل والمتغير التابع في فرض يمكن بناؤه؟

2- بين من خلال دراسة سسيولوجية لبعض متغيرات التحكم؟

3- ماذا تعني بـِ:

- المجتمع الإحصائي؟
- عناصر المجتمع الإحصائي؟

1- أعطِ مثالاً يبين العلاقة بين متغيرين مع التحكم في متغير ثالث؟

2- بكلماتك الخاصة، بيّن العلاقة بين الإحصاء وعملية تصميم البحث الاجتماعي؟

3- المطلوب مراجعة أحد المجالات العلمية في مجال العلوم الاجتماعية، ثم قم باختيار بحث في مجال اهتمامك مبيناً:

- ما الإحصاءات التي تم التركيز عليها في هذه الدراسة؟
- هل هذه الدراسة قد أسست على عينة من مجتمع إحصائي؟
- كم حجم العينة؟ كيف تم اختيار هذه العينة؟
- هل النتائج يمكن تعميمها على بعض المجتمعات الإحصائية؟
- ما هي المتغيرات التي تم استخدامها في هذا البحث؟
- بيّن المتغير المستقل والمتغير التابع بهذه الدراسة؟
- حدّد مستوى القياس لكل من هذين المتغيرين؟
- ما التقنيات الإحصائية التي تم استخدامها في هذه الدراسة؟
- حاول أن تتبع التحليل الإحصائي لكي تعرف مدى فهمك لهذه التحليلات؟

الهوامش والمصادر:

أولا: الهوامش:

1- George DiEKHoff, Statistics for the Social and Behavioral Sciences: Univariate, Bivariate, Multivarite, Mc Graw Hill, INC, 1992, P.4.

2- لزيادة التوضيح يمكن الرجوع إلى: عبد الله عامر الهمالي، أسلوب البحث الاجتماعي وتقنياته، منشورات جامعة قاريونس، 2003 م.

3- J. Richard Kendrick, Social Statistics: An Introduction using SPSS for Windows Pearson Education, INC. USA, 2005, PP. 15 - 16.

4- عبد الله عامر الهمالي، أسلوب البحث الاجتماعي وتقنياته، ص ص 210 – 212.

5- J. Richard Kendrick, Social Statatics, USA, 2005 P.16.

6- عبد الله عامر الهمالي، التقنيات الإحصائية ومناهج البحث، منشورات جامعة قاريونس، 2008، ص 282.

ثانياً: المصادر:

1- Colin Robson, Real World Research, Black well Publishing, USA, 2002.

2- George DiEKHoff, Statistics for the Social and Behavioral Sciences: Univariate, Bivariate, Multivarite, Mc Graw Hill, INC, 1992.

3- Joseph F. Healey, TheEssentials of Statistics: A Tool for Social Research, Wadsworth Cengage Learning, USA. 2010.

4- J. Richard Kendrick, Social Statistics: An Introduction using SPSS for Windows Pearson Education, INC. USA, 2005.

الفصل الثاني

المتغيرات والقياس

تهتم العلوم الاجتماعية والسلوكية بالاستقصاء غير المتناهي من التساؤلات في مجالات واسعة من حقول المعرفة كالاقتصاد والاجتماع والعلوم السياسية وعلم النفس إلى آخره من حقول المعرفة في مجال العلم الاجتماعي.

وعادةً ما تطرح هذه الحقول المعرفية العديد من التساؤلات مثل:

أ- ما هو دخل الأسرة في حي معين؟

ب- ما هو معدل الجريمة؟

ج- ما هي التغيرات التي حدثت في المستوى التعليمي خلال العقود الأربعة في المجتمع العربي الليبي؟

د- ما هي اتجاهات الناس حول تولي المرأة المهام الإدارية العليا؟

وبالرغم من الاختلاف في الأسئلة المطروحة في هذا الحقل أو ذاك، إلا أن القاسم المشترك بين هذه الحقول العلمية يكمن في كونها تركز على الاستقصاء المعرفي لواحد أو أكثر من متغير. فالمتغير: "عبارة عن خاصية تجريبية تتخذ قيمتين أو أكثر فإذا كانت هذه الخاصية قابلة للتغير كماً أو نوعاً فإننا ننظر إليها كمتغير"[1]. فعلى سبيل المثال، العمر يكون متغيراً حينما يمكننا بيان الفرق العمري بين شخص وآخر.

وتجدر الإشارة هنا إلى أن الفكرة المقابلة للمتغير هي الثبات "Constant" ونعني بالثبات حالة أو خاصية لا تباين فيها بين الحالات. فعدد الدراهم في ليبيا ثابت – أي أنّ كل دينار يساوي ألف درهم، وعلى أية حال فإن معظم الأبحاث تكرس جهودها لفهم تلك المتغيرات، ولماذا يتخذ هذا المتغير خاصيات معينة لبعض الحالات دون الحالات الأخرى. إن المهمة الأساسية لتسجيل الكيفية التي يظهر بها المتغير عبر مجموعة من الحالات يطلق عليها عملية القياس أو الملاحظة [2].

فالقياس هو العملية التي من خلالها يستطيع الباحث أن يحدد ويسجل الخاصيات الممكنة للمتغير كحالة مفردة. وبمعنى آخر، فالقياس هو طريقة خاصة تتبع في قياس المتغيرات والمفاهيم الاجتماعية فمتغير النوع يأخذ قيمتين أو سمتين (ذكراً أو أنثى) ويحدد لنا القياس ما هي الفئة التي يقع فيها هذا الشخص.

إن هذه المشاهدات والقياسات للمتغير في حالة البيانات الخام المتعلقة بالبحث والتي قد نختارها من وحدات التحليل يطلق عليها حالات؛ فالحالة هي خاصية تعكس عملية القياس المرتبطة بالمتغير، فعلى سبيل المثال، إذا ما أردنا أن ندرس عملية معدلات التذكر لدى طلاب المدارس الثانوية في منطقة محددة، فالحالات المدروسة هنا هي تلك المدارس الثانوية التي مُهرت بعلامات عالية تشير إلى معدل التذكر المطبق على هذه المدارس.

في هذا المثال فإن قائمة كل المدارس الثانوية في تلك المنطقة تمثل المجتمع المستهدف، فالمجتمع هو مجموع الحالات الخاضعة لدراسة في حين تعني كلمة مجتمع في الحياة اليومية مجموع السكان أو الناس الذين يعيشون في بلد معين أو منطقة معينة، إلا أنه يمكننا الإشارة إلى أن الباحث ليس في مقدوره دراسة أفراد المجتمع المستهدف بكامله وبالتالي يكون لديه حرية اختيار جزء من هذا المجتمع وهذا الجزء هو ما يطلق عليه العينة.

فالعينة تعني مجموعة الحالات التي لا تحتوي على كل الحالات السكانية وإنما جزء من السكان. فعلى سبيل المثال، أنه من الصعوبة بمكان أن نكون قادرين على استقصاء كل المدارس في المنطقة وبالتالي فقد يلجأ الباحث لاختيار عشر مدارس لاستقصاء وقياس معدلات التذكر لهذه العينة.

ولتلخيص المفاهيم الأساسية التي أوردناها فيما سبق (المتغير، الحالات، القياس، المجتمع، العينة) انظر جدول رقم (1).

دعنا ننظر إلى واحد من الأمثلة التي تم التطرق إليها في بداية هذا الفصل والذي يمثل السؤال البحثي التالي: ما هو دخل الأسرة في حي معين؟

جدول (1): ملخص

• المتغير	الدخل.
• الحالات	الأسر.
• القياس	تحديد الدخل لهذه الأسر بعينها.
• المجتمع	كل الأسر في المنطقة عند تاريخ معين.
• العينة	من مجموع العائلات في المنطقة التي اختيرت وتم قياس الدخل من خلالها.

وبعد أن تناولنا هذه المفاهيم من خلال السؤال الذي تم طرحه الآن يمكننا أن نواجه العملية البحثية الفعلية وهي كيف يمكننا قياس دخل الأسرة؟ وما هي الأداة التي يمكننا استخدامها لكي نحدد التباين بين الأسر فيما يتعلق بالدخل؟

إنه من أجل قياس متغير لعدد من الحالات فإن التحدي الكبير الذي يواجهنا يكمن في عملية التعريفات التصورية والتعريفات الإجرائية.

التعريفات التصورية والإجرائية للمتغيرات [2]:

عند الحديث عن التعريفات التصورية والإجرائية للمتغيرات يمكننا طرح الأسئلة التالية: من أين جاءت هذه المتغيرات؟ ولماذا اختار الباحث دراسة هذه المتغيرات بشكل خاص دون غيرها؟

ولنجيب على هذه الأسئلة فإننا نود الإشارة إلى أن اختيار المتغيرات الخاضعة

للاستقصاء عادة ما تتأثر بعدد كبير ومعقد من العوامل. ويمكننا في هذا السياق أن نركز على ثلاثة عوامل وهي كالتالي:

1- الإطار النظري:

عادة ما توجد عدة طرق نظرية لتفسير الواقع من حولنا وغالباً ما تُؤْخَذ هذه التفسيرات دون جدال أي كحقيقة ثابتة بأن هذا المتغير جدير بالدراسة والاهتمام.

وتجدر الإشارة هنا إلى أنه يمكننا أن نتعامل من خلال نظرية تقليدية قائمة تهتم بمتغيرات معينة وتمثل هذه المتغيرات الأساس في نظرة هذه النظرية للعالم الواقعي. فعلى سبيل المثال حيث تركز النظرية الماركسية على الطبقة الاقتصادية كمتغير أساسي ومهم للبحث، بينما اتجاه نظري آخر قد لا يعتمد مثل هذا المتغير ويعتبره شيئاً غير ذي جدوى.

إن تحليل العالم الواقعي من خلال الطبقة الاقتصادية يعني أننا لا يمكن أن نحلله إلا في هذا الإطار دون غيره، ومثل هذه الممارسة لا يمكن أن نطلق عليها وصف ممارسة سيئة أو جيدة. فبدون نظرية – من أجل أن يكون لدينا تصور واضح حول العالم الواقعي– تصبح المشاهدات المتعلقة بالبحث غير مرتبطة ببعضها البعض بطريقة علمية ومنطقية.

وتجدر الإشارة هنا أيضاً إلى أن التصورات المسبقة التي على أساسها تم اختيار المتغيرات للاستقصاء غالباً ما تكون الأساس العلمي للبحث.

2- الأجندة المحددة مسبقاً للبحث:

في بعض الأحيان قد لا يستطيع الباحثون أنفسهم تحديد السؤال البحثي والمتغيرات المطلوب دراستها أو استقصاءها، فعلى سبيل المثال قد تطلب جهة رسمية التعاقد مع بعض الباحثين لإجراء دراسة معينة أو بحث تم تحديده سلفاً من قِبل الجهة الممولة، في مثل هذا الموقف فإن الشخص أو الأشخاص يقومون بإجراء البحث بالرغم من وجود مساحة قليلة لاختيار المتغيرات المزمع استقصاءها والكيفية التي يمكن بها تعريف هذه المتغيرات باعتبار أن الباحثين يقومون ببحث لأشخاص آخرين أو جهات خاصة.

3- حب الاستطلاع يقود للبحث:

في بعض الأحيان قد لا يكون لدينا تعريف واضح للإطار النظري للاعتماد عليه في الدراسة، وكذلك لا يوجد لدينا توجيه واضح من شخص أو جهات معينة للمفاهيم الرئيسية المزمع استقصاءها وعوضاً عن ذلك فنحن نريد أن نستقصي متغيراً على أساس الاستطلاع أو على أساس تصور شعوري غير قوي قد يكون في بعض الأحيان مهماً لنشكل من خلاله متغيراً معيناً. إن مثل هذه الأحوال تكون سبباً مهماً للقيام بالبحث كنظريات أمرية.

حقاً عندما تتحول إلى حقل كلي للبحث وخاصة عندما توجد نظريات مجازفة فببساطة يمكننا الاندفاع إلى بحث ناضج جداً.

إن هذه الدوافع الثلاثة واضحة بجلاء فهي مانعة التبادل، فعلى سبيل المثال حتى وإن حددت أجندة البحث من قِبل جهات محددة فإن هذه الجهات سوف تتعامل وبكل تأكيد مع هذه الأجندة البحثية من خلال بعض الإطارات النظرية، ومهما كان الدافع فإن الاستقصاء الاجتماعي أساساً يقودنا إلى متغيرات محددة للاستقصاء. تكمن المرحلة الأولية لأي متغير في تعريفه تعريفاً تصورياً فالتعريف التصوري أو (التعريف الاسمي) لمتغير عادة ما يستخدم مصطلحات حرفية لتحديد خاصيات المتغير.

والتعريف التصوري يشبه التعريف القاموسي؛ حيث يمدنا بتعريف عامل للمتغير حتى يمكننا أن نتحصل على إحساس عام حول ما يعنيه هذا المتغير، فعلى سبيل المثال إذا ما أردنا قياس الدخل فإنه بالإمكان أن نعرف الدخل تصورياً بأنه المطلب الشرعي لشخص ما في الحصول على السلع والخدمات[3].

إنه من الوضوح إذا ما أردنا توجيه الباحثين لقياس شرعية المطالبة لشخص ما للسلع والخدمات فإنهم سيكونون في حيرة من أمرهم ولذلك فإنه يتوجب علينا تقديم مجموعة من التوجهات تسمح للباحثين فعلياً تسجيل كم من المطالب على السلع والخدمات التي سوف تتباين من شخص إلى شخص آخر، بمعنى آخر، فإن أول مرحلة لتعريف المتغير المطلوب دراسته تكمن في تعريفه على المستوى التصوري كما نحتاج إلى القواعد

والإجراءات - عمليات إجرائية - التي تسمح فعلياً بملاحظة المتغير في العالم الواقعي، والسؤال المطروح هو: ما الشيء الذي نبحث عنه لكي نبين دخل شخص ما؟ إن طرح مثل هذا السؤال يقودنا إلى مشكلة التعريفات الإجرائية، فالتعريف الإجرائي لمتغير معين يحدد الإجراءات والمحكّات التي تؤخذ لقياس ما يحتويه هذا المتغير لمجموعة من الحالات، فإذا ما أردنا مشاهدة دخل شخص ما فإن الحاجة تتطلب أن نقرر ما هي الأشياء التي ننظر إليها لكي تساعدنا لقياس هذا المتغير.

إن بيان الدخل هو مجموع كل ما يتحصل عليه الفرد بشكل مباشر مثل الأجور والمعاشات والمساعدات الاجتماعية خلال العام المنصرم، مثل هذه الأحوال هي الأساس لتعريف الدخل وبهذا التعريف الذي تَحَصَّل عليه الباحث يمكنه أن يذهب إلى العالم الواقعي ويبدأ في قياس دخول الأفراد مضافاً إليه كل النقود التي تحصّل عليها الفرد من مختلف المصادر في العام السابق.

وتعد عملية التعريف الإجرائي هي الأساس وإذا لم تكن هذه العملية هي الأساس فإن مصادر التباين سوف تحدث في البحث.

إن أي تعريف تصوري معطى باستطاعتنا تعريفه تعريفاً إجرائياً بعدة طرق مختلفة ولا يمكن النظر لواحد من هذه الطرق على أنه تعريف مطلق أو تام، على سبيل المثال، إن التعريف الإجرائي للدخل من خلال مجموع ما يتحصل عليه الشخص بشكل مباشر يترك هذا التعريف بعض المصادر الأخرى للدخل مثل الحصول على الخدمات النوعية بدلاً من الدفع النقدي المباشر لشخص.

والسؤال الذي يمكن طرحه هنا هو: ما هي المحكّات التي يمكن أن نستخدمها لكي نقرر ما إذا كان التعريف الإجرائي دقيقاً أم لا؟ إن المحك الذي يمكن استخدامه لمعرفة ما إذا كان التعريف الإجرائي صحيحاً أم لا؟ يُعْرَف في التقنيات الأدبية بمشكلة صدق المفهوم.

تصورياً يمكننا النظر إلى التعريف الإجرائي بأنه يتباين تحت المتغير الذي نفكر فيه، فميزان الحرارة الزئبقي يعد أداة جيدة لقياس التغيرات اليومية في درجة الحرارة وعندما

نتعامل مع التغيرات في المتغير "درجة الحرارة"، في وسيلة قياسه "على المزلاج الزئبقي" نجده يتغير، نفترض أننا قمنا بملء ميزان الحرارة الزئبقي بالماء بدلاً من الزئبق فإن التباينات في درجة الحرارة اليومية لن تتناسب مع التغيرات في ميزان الحرارة، وخلال يومين ربما في الحقيقة يكون هناك اختلاف في الحرارة بدون أن يكون هناك تباين يمكن أن يسجل بواسطة الأداة.

وبالرجوع إلى المثال السابق المتعلق بالدخل وبالاعتماد على التعريف الإجرائي والذي تضّمن الدفع النقدي المباشر فإنه يمكننا أن نسجل شخصين لديهما نفس الدخل وإن كانا في الحقيقة مختلفين. تصور أن شخصين قد استلما نفس كمية النقود مقابل العمل ولكن أحدهما لديه أطفال تم دفع مصاريفهم الدراسية من خلال المؤسسة التي يشتغل بها وبوضوح فإنه لا يوجد تباين بين هذين الشخصين فيما يتعلق بالدخل بالنظر إلى سيطرتهم على السلع والخدمات ولكن هذا التباين لن يسجل إذا ما اعتمدنا على التعريف الإجرائي القائم فقط على مجموع الدخل النقدي المباشر.

ولزيادة التوضيح فإذا ما انتقلنا من التعريف التصوري إلى التعريف الإجرائي لمتغير، يتوجب علينا أن نضع في الاعتبار المثال التالي:

إذا ما رغب الباحث في دراسة الجريمة فإنه يمكنه أن يعرف السلوك الإجرامي تصورياً بأنه أفعال غير قانونية من العنف ضد أعضاء آخرين من المجتمع أو ضد ممتلكاتهم، فالسؤال المطروح هنا لدى الباحث هو: كيف يمكنه أن يبين أو يعين نمط التباين في هذا المتغير؟

توجد لدينا مجموعة من التعريفات الإجرائية يمكن اعتمادها:

أ- حساب عدد المرات التي تم فيها العنف الإجرامي من السجلات الرسمية.

ب- حساب كمية الوقت التي قضاها الشخص في السجن.

ج- طرح مجموعة من الأسئلة على المبحوثين لمعرفة ما إذا ارتكبوا أية جرائم.

د- تسجيل لون شعر الرأس لدى الفرد.

وبوضوح سوف يكون من الصعوبة إثبات إن التعريف الإجرائي الأخير صحيح. إنه من غير الممكن أن نقول أن مستوى السلوك الإجرامي لدى الفرد يتغير بتغير لون شعر رأسه !.

إن التعريفات الإجرائية الأخرى تبدو أكثر التصاقاً إلى المفهوم العام للسلوك الإجرامي إلا أن كل واحد من هذه التعريفات لديه مشكلته الخاصة. فعند طرح السؤال على الأشخاص إذا ما ارتكبوا أي جريمة ربما لا يعكس هذا السؤال المقياس الصحيح لأن الناس قد لا يكونون صادقين في الإجابة حول السؤال لحساسيته أما فيما يتعلق بالسؤال حول عدد الأوقات التي تم فيها القبض على الشخص يكون غير صحيح. فقد يكون هناك شخصان في الحقيقة لديهم نفس مستوى السلوك الإجرامي إلا أن أحدهما قد يكون عدد حالات القبض المسجلة عليه أكثر بسبب انتمائه إلى الجماعات المهمشة والتي غالباً ما يكون الهدف الأساسي لدى الشرطة للقبض عليه.

إن هذا التعريف في الحقيقة يتطلب قياساً متغيراً مختلفاً عن ذلك المتغير الذي نهدف إليه. إن تحيز الشرطة أكثر إلى السلوك الإجرامي وباستخدام أيٍّ من هذه التعريفات الإجرائية لقياس إجرامية الشخص قد لا تكون مرآة صادقة لمعرفة السلوك الإجرامي على حقيقته[4].

هناك مجموعة من العوامل يمكن أن تولّد مشكلات للوصول إلى التعريف الإجرائي لمتغير بدرجة عالية من صدق المفهوم.

1- تعقد المفهوم:

إن بعض المتغيرات قد لا تكون معقدة فنوع الشخص على سبيل المثال يمكن تحديده بخاصيات فيزيقية وعلى أي حال فإن معظم المتغيرات نادراً ما تكون واضحة فقد وجدنا متغير الدخل يحتوي على جملة من الأبعاد مثل: الدفع النقدي المباشر والدفع غير المباشر أي (النوعي) وأن كل بعد من أبعاد الدخل هي متغيرات تصورية في ذاتها وتطرح مجموعة من المشكلات الإجرائية المتعلقة بها وأن كل واحد من هذه التعريفات يركز على واحد من هذه الأبعاد فالتركيز على الدفع النقدي على سبيل المثال يستبعد الأبعاد الأخرى.

2- توفر البيانات:

قد نتمكن من الحصول على تعريف إجرائي ليساعدنا على تحديد المتغير الذي نرغب في دراسته بشكل كامل، فعلى سبيل المثال يمكننا أن نفكر في عدد من حالات القبض بطريقة متصدعة لمشاهدة السلوك الإجرامي، فالباحثون قد لا يسمح لهم – نتيجة لسرية المعلومات – بمراجعة سجلات الشرطة للحصول على المعلومات المطلوبة، وعليه فإنه من غير الواضح الحصول على تعريف كامل للتعريفات الإجرائية التي يمكن استخدامها وذلك لعدم قدرتنا في الحصول على بيانات مثالية.

3- تكلفة وصعوبة الحصول على البيانات:

ولنقل إننا قادرين على مراجعة سجلات الشرطة والحصول على عدد الحالات المقبوض عليها، إن التكلفة في الحصول على ذلك يمكن أن تكون غير مسموح بها من حيث الوقت والمال وبنفس الكيفية يمكننا الشعور بأن قياساً محدداً لتلوث الماء هو مثال لتقويم تآكل النهر ولكننا نحتاج للاستعانة بخبير بتقنيات قياسية عالية بإمكانها إعاقة هذا التآكل كخيار، بدلاً من اللجوء إلى قرار مزاجي بأن الماء ضبابي يمكن أن يكون مفضلاً كقياس سريع وسهل.

4- الأخلاق:

قد تسمح لنا الشرطة بالاطلاع على المعلومات الخاصة بالقبض على الأفراد كما قد يكون لدينا المال والوقت الكافي، فهل هذا يعطينا المبرر للاطلاع على الوثائق التي ليست جزءاً من مشروع البحث؟ إن مسألة الأخلاقيات – معرفة الصواب من الخطأ – هي قضية شائكة وقد لا نستطيع إثارتها بشكل جدي هنا وإنما بطرحها كمشكلة تؤثر في التعريف الإجرائي للمتغيرات التي تظهر بشكل انتظامي في الواقع الاجتماعي المتعلق بحياة الناس.

ولهذه الأسباب أو تلك هناك جدال واسع حول مسألة الصدق Validity التي تتعلق بمشكلة التعريفات الإجرائية، وفي الحقيقة فإن هذه الجدالات تدور حول البحوث

الكمية وليست في واقع الأمر حول التقنيات أو نتائج البحث ولكن السؤال الذي يمكن طرحه هو فيما إذا كانت المتغيرات قد تم تعريفها بشكل صحيح في المقام الأول.

وتجدر الإشارة إلى أنه إذا لم نستخدم المحك الإجرائي لقياس متغير يكون ذا حساسية في الطريقة التي يتغير بها المتغير بين الحالات قد يقودنا إلى توليد نتائج غير مرضية.

فالتعريف الإجرائي لمتغير سوف يكون مألوفاً في تحديد مدى الفئات أو القيم "في بعض الأحيان نطلق عليها الدرجات" لمتغير من خلال الحالات الفردية التي يحتويها.

ونعني بهذه الفئات والقيم الحصول على مدى تباين الاحتمال الذي قد يظهر أثناء عملية القياس، فالفئات المتعلقة بالذكور والإناث يمكن تحديدها على سبيل المثال، إن المدى الكلي لتباين الحالات الفردية يمكن أن يظهر في متغير النوع وبتحديد ما لهذا المدى نكون قد أوفينا بالقواعد المتعلقة بالتعريف الإجرائي، فالتعريف الإجرائي يساعد الباحث في تحديد كل حالة في الفئة المخصصة لها في المتغير.

إن ما تناولناه في هذا الإطار هو في الحقيقة يحتوي على عنصرين أساسيين منفصلين من القياس:

أولهما: ما يطلق عليه أسس مانعة التبادل والتي تشير إلى أنه لا يمكن أن تحتوي أي حالة أكثر من قيمة واحدة لنفس المتغير، فعلى سبيل المثال، قد لا يكون عمر شخص ما ثماني عشرة سنةً، وأربعاً وستين سنة.

ثانيهما: أن القياس يجب أن يتبع أسس الشمولية التي تشير إلى أن كل حالة يمكن تصنيفها إلى مجموعة فئات، فعلى سبيل المثال، فإن المقياس لقياس الوضع العائلي يجب أن يتبع كل نمط احتمالي للوضع العائلي الذي يمكن أن يظهر وإذا لم نشمل فئة " لم يتزوج على الإطلاق " فإن هذه الحالة سوف تمدنا بتصدع خلال بعض الحالات التي تقع والتي لم نبينها في القياس[5].

تصنيف المتغيرات وفقاً لمستوى القياس:

مستويات القياس:

التعريف الإجرائي للمتغير: هو تحديد أو تعيين مدى الفئات أو القيم التي يمكن تخصيصها للحالات الفردية التي تتضمن مستوىً معيناً من القياس.

وعليه توجد لدينا أربعة مستويات للقياس وهي:

المستوى الاسمي والمستوى الترتيبي والمستوى ذو المسافات والمستوى النسبي.

وتمثل هذه المقاييس الأربعة الأساس التميزي في الإحصاء والذي من خلاله نستطيع تحديد كم من المعلومات يمكن جمعها.

في الحقيقة عندما نقرر ما هي التقنيات الإحصائية المتوفرة والتي يمكننا أن نعتمدها في تحليل البيانات فإن السؤال الذي يتبادر إلى الذهن هو: ما هو المستوى الذي على ضوئه يمكن قياس المتغير؟ وكما سنرى، هناك أشياء يمكن عملها للبيانات التي تم جمعها على مستوى المقياس ذي المسافات قد لا نستطيع عملها مع بيانات ثم جمعها على المستوى الاسمي فكلما كان المستوى من القياس عالياً كانت المعلومات المتعلقة بالمتغير كثيرة.

وفيما يلي عرضاً مفصلاً لهذه المقاييس:

1- المقياس الاسمي:

يعتبر المقياس الاسمي أدنى المقاييس ويشير هذا المقياس إلى الأعداد والرموز ويعتمد على النظام التصنيفي للأشياء أو الأشخاص، فعلى سبيل المثال نفترض أننا نرغب في معرفة الديانة للأفراد فإنه يمكننا إجرائياً تعريف ديانة لشخص ما بالرجوع إلى انتماءاته الدينية والتي يمكن تحديدها من خلال مدى الفئات التالية:

الإسلام – الهندوسية – اليهودية – المسيحية..... إلى آخره.

وتجدر الإشارة إلى أنه لضمان المقياس صحيحاً فإنه لابد أن يكون شاملاً وذلك باحتوائه على كل الفئات، إلا أن احتواء الفئات يمكن أن يكون خادعاً، وهناك طريقة

أخرى بسيطة تساعدنا على كشف المقياس الاسمي في كونه صحيحاً وذلك بإعادة ترتيب الفئات لنرى ما إذا كان الترتيب يحتفظ بمنطقيته، فعلى سبيل المثال إن أحد الترتيبات التالية لقائمة الديانات يكون صحيحاً:

المسيحية	الإسلام
الإسلام	اليهودية
اليهودية	الهندوسية
الهندوسية	المسيحية
أخرى	أخرى

وبوضوح فإن هذا الترتيب الذي ظهرت به هذه الفئات لم يتأثر بإعادة التصنيف لأن كل هذين التصنيفين لازالا يحتفظان بقاعدة مانعة التبادل والشمولية.

إن المتغير الذي يمكن قياسه على المستوى الاسمي يتباين كيفياً وليس كمياً لأن الشخص ذا الديانة المسيحية يختلف عن شخص آخر في فئة الديانة الهندوسية بالنظر إلى متغير الدين لكنهما لا يمتلكان أكثر أو أقل دين.

إن متغير الدين يمكن ترميزه بالشكل التالي:

1- الإسلام.

2- اليهودية.

3- الهندوسية.

4- المسيحية.

5- أخرى.

إن هذه الترميزات للفئات لا تحتوي على أي معنى كمي بالشكل المعتاد ولكن الأرقام التي أعطيت كان الهدف منها ببساطة هو التعريف بالفئات المختلفة ولكن ذلك لا يعبر عن العلاقة بين هذه الفئات.

تجدر الإشارة إلى أنه بإمكاننا وبكل بساطة أن نستخدم الترميز التالي لتعيين القيم الرقمية لكل فئة:

1- الإسلام.
2- اليهودية.
3- الهندوسية.
4- المسيحية.
5- أخرى تذكر..

2- المقياس الترتيبي:

بالإضافة إلى خصائصه كمقياس ترتيبي فإنه يمتلك خاصية التصنيف وكذلك فئاته مانعة التبادل وشاملة. حيث ترتب فئات هذا المقياس على حسب الخاصية لكل فئة "الترتيب الأول - الترتيب الثاني - الترتيب الثالث، الأكبر - الكبير - الأصغر..... إلى آخره".

وتساعدنا المقاييس الترتيبية في ترتيب الحالات ويتضمن هذا الترتيب ترتيب الحالات بمعنى كمي مثل: من الأدنى إلى الأعلى ومن القليل إلى الكثير أو من الأضعف إلى الأقوى ويشاع استخدام المقاييس ترتيبياً عندما يتعلق الأمر بقياس الاتجاه أو الرضا في مسوح الرأي، فعلى سبيل المثال نفترض أننا نحاول قياس الدخل فإننا يمكننا أن نحدده على المقياس التالي:

دخل منخفض - دخل متوسط - دخل عالٍ.

وتحتوي فئة ذوي الدخل المنخفض أولئك الذين يصل دخلهم إلى 15.000 ألف دينار أو أقل في حين تشمل فئة ذوي الدخل المتوسط أولئك الذين يكسبون ما بين 15.001 دينار و 50.000 ألف دينار في السنة في حين يصل ما يكسبه ذوي الفئات العليا إلى أكثر من 50.000 ألف دينار في السنة.

إن ترتيب مثل هذه الفئات ليست من مهام المقياس الاسمي ذلك المقياس الذي من خلاله يمكن تعيين الحالات إلى فئات.

وبالإضافة إلى خصائص المقياس الترتيبي في كونه يرتب الحالات فإنه يساعدنا أيضاً في معرفة شخص ما ينتمي إلى فئة ذوي الدخل المتوسط وأن لديه دخلاً أكثر من الشخص الذي يقع في فئة ذوي الدخل المنخفض.

وبطريقة أخرى فإن الشخص الذي يقع في فئة الدخل المتوسط يمكن وضعه في الترتيب الأعلى من الشخص الذي ينتمي إلى ذوي الدخل المنخفض.

إن الاختلاف بين بيانات المقياس الترتيبي وبيانات المقياس الاسمي يتمثل في كون أن الحالة في المقياس الاسمي لا تختلف من حالة إلى أخرى من حيث كونها أفضل أو أقوى أو أكبر أو أكثر قوة، فبالنظر إلى المقياس التالي:

(دخل منخفض - دخل متوسط - دخل مرتفع) نـجده يحتفظ بخصائص المقياس الترتيبي وذلك باحتفاظه بالترتيب المنطقي بين الفئات، فإذا ما قمنا بإعادة ترتيب هذه الفئات حسب خصائص المقياس الاسمي فإن ترتيب الحالات طبقاً للدخل قد يُفْقَد.

دخل متوسط - دخل عالٍ - دخل منخفض

وكما رأينا في بيانات المقياس الاسمي فإن القيم العددية يمكن أن تخصص للفئات كشكل من أشكال الاختزال في حين تصبح هذه القيم العددية في حالة المقاييس الترتيبية ضرورة للحفاظ على منطقية الترتيب وعليه فإن أحد الطريقتين لمجموعة الأرقام التالية يمكن استخدامها:

(1)	(2)	(3)
الدخل المنخفض	الدخل المتوسط	الدخل العالي
(23)	(88)	(105)
الدخل المنخفض	الدخل المتوسط	الدخل العالي

وتجدر الإشارة هنا إلى أنه يمكننا القول إلى أيٍ من نظامَيْ الترميز السابق يسمح للفئات بأن تعرف وترتب بالنسبة لبعضها البعض إلا أن هذه الأرقام في حد ذاتها ليس لديها أي دلالة كمية أكثر من كون أن وظيفتها هي الترتيب.

3- مقياس ذو المسافات والنسبي:

إذا كان المقياس الترتيبي يسمح لنا بترتيب الحالات فيما يتعلق بمتغير ما بالتالي يمكننا القول بأن شخصاً ما أو حالة ما أفضل أو أكثر قوة من الحالة الأخرى فإنه لا يسمح لنا بمعرفة الاختلاف الحقيقي بين الحالات كذلك لا يدل على مدى أو مقدار ما تمتلكه كل حالة من هذه الحالات.

وإذا ما استخدمنا المثال السابق لقياس الدخل فإننا لا نستطيع القول: كم أكثر فعلياً يكسبه شخص ما في فئة ذوي الدخل العالي مقارنة بشخص ما في فئة ذوي الدخل المتوسط، إنه قد يكون من غير السليم لنا أن نستخدم خطة الترميز الثانية المشار إليها في أعلاه، ونقول إن الشخص الذي يتحصل على دخل عالٍ يزيد بـِ 17 وحدة من دخل الشخص في الفئة ذات الدخل المتوسط، على سبيل المثال 105 – 88 – 17، ففاصل المسافة بين الفئات غير معروفة.

مثال: نفترض أننا نريد أن نقيس الدخل بطريقة أخرى وذلك من خلال طرح السؤال التالي على كل شخص لمعرفة دخل كل منهم في السنة السابقة "بالدينار الليبي" وبشكل واضح فإنه بإمكاننا أن نحدد لكل شخص الفئة التي يندرج تحتها ذلك بناءً على المعلومات التي تحصلنا عليها "القيمة المالية السنوية بالدينار" كذلك يمكننا ترتيب هؤلاء الأشخاص طبقاً لهذا المقياس "ذي المسافات" وذلك من خلال بيان أن شخصاً ما يمتلك أكثر أو أقل دخل من شخص آخر كذلك يمكننا قياس كمية الفرق في الدخل بين الحالات حيث يختلف المقياس ذو المسافات عن المقياس الاسمي والترتيبي في كونه يمتلك كمية الفرق في الدخل بين الحالات كما أن المقياس يمكّنُنا من أن نتحصل على الأعداد التي تعكس حقيقة القيم الكمية، أي مجموع الدينارات.

وتكمن قدرة هذا المقياس في قياس المسافات بين النقاط على المقياس ويختلف المقياس ذو المسافات عن المقياس النسبي من حيث أن المقياس النسبي يمتلك قيم الصفر المطلق التي تشير إلى عدم وجود كمية المتغير.

ومجمل القول فإنه من خلال المقياس ذي المسافات والمقياس النسبي يكون في مقدور

الباحث ليس فقط معرفة أن حالةً ما تمتلك أكبر أو أقل للمتغير تحت الدراسة من الحالة الأخرى ولكنه يستطيع القول كم هي أكثر أو أقل، فعلى سبيل المثال: إذا كان شخص ما دخله 20.000 ألف دينار ليبي فهو يمتلك أكثر بـِ 10.000 دينار ليبي من الشخص الذي يمتلك دخلاً يساوي 10.000 دينار ليبي في السنة وبهذا نستطيع أن نتعرف على المسافة بينهما، بل أكثر من ذلك فإن المسافة بين النقاط على المقياس هي قيم متساوية على المستوى العام للمدى فالفرق بين 20.000 ألف دينار ليبي و 10.000 دينار ليبي هو نفس الفرق في الدخل بين 120.000 ألف دينار ليبي و 130.000 ألف دينار ليبي.

وبوضوح فإن الأعداد أو الأرقام على المقياس ذي المسافات والنسبي ذات دلالة، حيث تشير هذه الأرقام إلى كمية قابلة للقياس. وعليه، فإن هذه الأرقام تبين قيم المتغير أو القيم المتعلقة بالمتغير. لاحظ أن "صفر" دينار يمثل حالة لا تمتلك كمية متغير الدخل وتعرف هذه الحالة بنقطة الصفر الحقيقية. وتشير إلى خاصية البيانات النسبية كنقيض للبيانات ذات المسافات المتساوية. فعلى سبيل المثال، يمكن قياس الحرارة بدرجات مئوية لا تمتلك قيمة الصفر الحقيقي حيث لا توجد نقطة. إلا أن درجة الصفر المئوية لا تشير إلى حالة "حين لا توجد حرارة" فالجو بارد ولكنه ليس بهذه البرودة، وعوضاً عن ذلك، فإن درجة الصفر المئوي تشير إلى شيء آخر - النقطة التي يتجمد فيها الماء - وعليه، فإن هذا التمييز بين المقياس النسبي وذي المسافات، فيما يتعلق بالقياس هو أمر واضح في الذاكرة، وليس بالضرورة أن نتتبع ذلك التمييز. فبإمكاننا بشكل عام أن نجري نفس التحليل الإحصائي للبيانات المقاسة على مستوى ذي المسافات، والبيانات المقاسة على المستوى النسبي.

الفرق بين المتغيرات المقولية والمتغيرات العددية:

ولكي يقرر الباحث ما إذا كانت المتغيرات التي يتعامل معها متغيرات مقولية أو متغيرات عددية، ينبغي النظر إلى الكيفية التي على ضوئها يتم جمع البيانات. فعلى سبيل المثال، إن متغير العمر يمكن قياسه على أكثر من مستوى. كالسؤال الذي يمكن طرحه: أي من الفئات العمرية التالية تمثل عمرك؟ من هنا يكون متغير العمر متغير مقولي،

ومتغير عددي، لأن البيانات التي تم جمعها في التعداد السكاني على سبيل المثال، تمَّ على ضوئها حساب العمر. السؤال الذي ينبغي طرحه هنا هو كيف لنا أن نميز بين المتغيرات المقولية والمتغيرات العددية؟

1- المتغيرات المقولية: الفرق بين المتغيرات الاسمية والترتيبية:

المتغيرات المقولية هي تلك المتغيرات التي تم جمع بياناتها من خلال فئات إجابة response categories تم تحديدها مسبقاً من قِبل الباحث.

ومن أمثلتها: المتغيرات الديموغرافية: النوع (ذكر / أنثى)، محل الإقامة (حضري / ريفي)، الحالة الزواجية (أعزب، متزوج، مطلق، أرمل)، والمتغيرات المتعلقة بالمواقف وعادة ما تجمع بيانات هذه المتغيرات في شكل فئات إجابة:

إلى أي حد يعتبر رأيك في أمور تتعلق بقرارات مهمة داخل أسرتك؟

- مهم جداً
- مهم إلى حد ما
- مهم إلى حد ما
- مهم قليلاً
- غير مهم

والمتغيرات المتعلقة بالسلوك تجمع بياناتها عادة في شكل فئات:

هناك بعض من الناس لديهم بعض الوقت للمساهمة في العمل التطوعي، بينما آخرون ليس لديهم الوقت. أنت شخصياً هل لديك الوقت للمشاركة في النشاطات التطوعية؟

- أكثر من مرة في الأسبوع.
- مرة واحدة في الأسبوع.
- بعض الأوقات شهرياً.
- مرة في الشهر.
- كل شهرين أو ثلاثة أشهر.
- على الأقل مرة في السنة.
- لا أشارك.

إن الفكرة وراء فئات الإجابة، هي أن تكون فئات الإجابة مانعة التبادل، بمعنى ألا تتداخل فئات الإجابة بحيث لا تتعدى أية إجابة الفئة المخصصة لها. قد نجد في بعض الأحيان بعض الأمثلة تتداخل فيها فئات الإجابة، مثل فئات العمر التالية:

18 – 25

25 – 35

35 – 45

+ 45 إلى آخره.

في مثل هذه الفئات يجد الباحث صعوبة في تحديد الفئة العمرية المناسبة لشخص أجاب بأن عمره 25 سنة أو 35 سنة. هذا لا يعني أن المبحوثين لن تكون لديهم صعوبة في تحديد الفئة المناسبة لمواقفهم أو سلوكهم أو أي خاصية أخرى تم قياسها. وعلى أية حال، هذا يعني أن خاصية واحدة متعلقة بالمبحوث يمكن أن تلائم أكثر من فئة للمتغير.

2- المتغيرات العددية:

يمثل المقياس ذو المسافات والنسبي المتغيرات العددية. أي تلك المتغيرات التي تجمع بياناتها كأعداد دون محاولة وضعها في فئات مسبقة من قِبل الباحث. ومن أمثلتها عدد السنوات التي قضاها شخص ما في التدريس الجامعي، عدد الساعات الأسبوعية التي يقضيها شخص ما في العملية التدريسية. لاحظ أن البيانات التي تم جمعها كمتغيرات عددية يمكن جمعها كمتغيرات مقولية (مئوية) أيضاً. ومن الأمثلة الشائعة في هذا السياق متغير الدخل، فبدلاً من أن يطرح السؤال على المبحوث: كم من النقود يتحصل عليها؟ يمكن جمع البيانات حول الدخل على المستوى العددي، كذلك يمكن للباحث أن يصنف هذه الإجابات في فئات:

أقل من 300 د.ل.

300 – 4900 د.ل.

5000 – 6900 د.ل.

7000 – 8900 د.ل.

9000 فما فوق.

المتغيرات المنفصلة والمتغيرات المتصلة:

تجدر الإشارة هنا إلى أن هناك تمييزاً آخر يؤثر في عملية القياس وهو التمييز بين المتغيرات المنفصلة والمتصلة، فالمتغير المنفصل: هو ذلك المتغير الذي لا تمكنه طبيعته من أن يأخذ جميع القيم بين حدي المتغير بل يتحرك عند أعداد معينة دون سواها فعلى سبيل المثال النوع متغير منفصل يمتلك فئتين احتماليتين هما ذكر - أنثى. ومقياس المتغيرات على المستوى ذي المسافات والنسبي فإن وحدة القياس عادة لا يمكن أن تكون قابلة للتقسيم فإذا ما نظرنا إلى عدد الأطفال في الأسرة فإنه من غير المنطق إن نقول أن لدى الأسرة 1.7 طفل باعتبار أنه لا يمكن أن يكون لنا أطفال بوحدة تقل عن طفل واحد.

وبالتالي يمكننا أن نقفز من القيمة الكلية للمتغير إلى الأخرى باعتبار أن المتغير المنفصل لا يحتوي على كسور. ولزيادة التوضيح نورد المثال التالي: لعدد السجناء لكل زنزانة، وعدد هيئات الرعايا في حي معين أيضاً عدد الحوادث الصناعية في السنة السابقة.

ويمكننا على الجانب الآخر أن نورد مثالاً متغير الرضا بالخدمات التي تقدمها المكتبة فإن مستويات الرضا بين مستخدمي المكتبة قد يكون بينها فروق ضئيلة ولذلك فإن مستخدمي المكتبة يمكن أن يتصوروا التغير بطريقة تدريجية من شخص لآخر أو لنفس الشخص عبر فترات زمنية متعاقبة هذا المثال ينطبق على المتغيرات المتصلة فالمتغير المتصل هو المتغير الذي تمكنه طبيعته من أن يأخذ جميع القيم بين حدي المتغير وإذا زاد رقم إلى الرقم التالي له يمر بكل الكسور الممكنة بينهما، فالعمر على سبيل المثال هو متغير متصل بحيث يمكن أن يقسم العمر إلى الشهور والشهور إلى أسابيع، والأسابيع إلى أيام... الخ.

إن الحد الوحيد هو بالضبط كيف نكون دقيقين فالسنوات ليست دقيقة كالشهور والشهور ليست دقيقة كالأسابيع.

نظرياً يمكننا من خلال المتغير المتصل أن نتحرك تدريجياً وبشكل سلس من واحدة من قيم المتغير إلى القيم الأخرى بدون عملية القفز.

عملياً نحن دائماً بإمكاننا القيام بعملية تقريب القياس والتعامل مع المتغير المتصل

كما لو كان متغيراً منفصلاً وهذا العلم يتيح للمقياس أن يقفز من قيمة إلى أخرى.. فعلى سبيل المثال يمكننا أن نرسخ على مقياس الرضا للخدمات التي تقدمها المكتبة بطرح أسئلة على الطلاب إذا ما كانوا راضين جداً أو راضين أو غير راضين أو غير راضين جداً من المقياس فهو مقياس منفصل بالرغم من أن المتغير في حقيقته متغير متصل.

وبنفس السياق ينطبق هذا الأمر على العمر فيمكننا قياس العمر بوحدات منفصلة مثل: السنوات أو الشهور، فالمتغير بطبيعته يزيد بطريقة متصلة.

إن استخدام المقاييس المنفصلة يُوفر لنا مجموعات منفصلة من الحالات العنقودية بمعنى آخر تعمل هذه المقاييس مثل مركز الجاذبية الذي يجذب كل التباينات الطفيفة القريبة منه في المتغير الذي يقلقنا أو يزعجنا.

فعندما نقول إن شخصين عمرهما ثمانية عشر عاماً فهما في حقيقة الأمر ربما يكونان مختلفين فيما يتعلق بالعمر إلا إذا ولدا بشكل دقيق في نفس الوقت إلا أن الفرق البسيط يمكن أن نراه بين شخص عمره ثمانية عشر عاماً وشهران وخمسة أيامٍ وساعتان وأثنتا عشرة ثانية، وشخص آخر عمره ثمانية عشر عاماً وثلاثة أشهر وأربعة عشر يوماً وسبع ساعات وثانية واحدة.

إن هذه القضية قد تكون لا علاقة لها بمشكلة البحث الذي نسعى من خلاله إلى استقصاء ومعالجة هذه الأعمار فيما يتعلق بهذا المتغير بالرغم من أنهما يختلفان حقاً [6].

إن الاختلاف البيّن بين بيانات المقياس الاسمي والترتيبي وذي المسافات والنسبي يكمن في كمية المعلومات التي يقدمها كل مستوى من هذه المستويات القياسية الأربعة.

جدول (2-1): الخصائص الأساسية لمستويات القياس

مستوى القياس	أمثلة	إجراءات القياس	العمليات الرياضية المسموح باستخدامها
الاسمى "أدنى المستويات"	النوع العرق الدين الوضع الزواجي	التصنيف إلى فئات	عدد الحالات في كل فئة مقارنة عدد الحالات في كل الفئات
الترتيبي	الطبقة الاجتماعية مقاييس الاتجاه والرأي	تصنيف الفئات + ترتيب الفئات في علاقتها ببعضها البعض	كل العمليات السابقة بالإضافة إلى الحكم بأكثر أو أقل من
ذو المسافات والنسبي "أعلى المستويات"	العمر بالسنوات عدد الأطفال الدخل	تصنيف الفئات + ترتيب الفئات إضافة إلى وصف المسافات بين الدرجات فيما يتعلق بتساوي الوحدات	كل العمليات السابقة بالإضافة إلى عمليات رياضية مثل الجمع، الطرح، الضرب... الخ

المصدر: Joseph F. HEALEY, The Essentials Statistics: A Tool for Social Research, Wadsworth Cengage Learning, USA, 2010, P.21.

يلخص لنا الجدول كمية المعلومات التي يزودنا بها كل مقياس من هذه المقاييس الأربعة، إن المهمة الأساسية تسمح لنا بالتعامل مع البيانات المجمعة على كل مقياس من هذه المقاييس فبيانات المقياس الاسمي تمتلك أقل المعلومات في حين بيانات المقياس الترتيبي تمتلك أكثر معلومات من المقياس السابق نظراً لإمكانية ترتيب الحالات في حين أن المقياس ذا المسافات والنسبي ينتزعان أكثر المعلومات لأنهما يساعدان في قياس الفرق بين القيم.

خلاصة:

حاولنا في هذا الفصل أن نناقش بشكل مبدئي الأحوال التي يجب أن نراعيها قبل استخدام المعلومات الإحصائية في التحليل، وقد تناولنا هذه القضية بشكل عام حيث إن باقي فصول هذا الكتاب ستتناول عملية التحليل وماذا نعمل بالبيانات التي يتم جمعها وبعد جمع البيانات يمكننا أن نخطو خطوة أخرى وهي عملية التحليل وأن أول خطوة في التحليل عادة ما تتعلق بوصف البيانات، وعملية وصف البيانات سنناقشها في موضعها في جزء آخر من هذا الكتاب.

كما تناولنا في هذا الفصل الفرق بين المتغيرات المقولية والمتغيرات العددية، وبشكل عام، فإن المتغيرات المقولية هي المتغيرات التي يتم جمعها من خلال فئات تم تحديدها مسبقاً من قِبل الباحث. في حين أن المتغيرات العددية هي تلك البيانات التي تم جمعها دون تحديد فئاتها مسبقاً. وتقسم المتغيرات المقولية إلى متغيرات اسمية ومتغيرات ترتيبية. فالمتغيرات الاسمية هي تلك المتغيرات التي لا تخضع فئاتها إلى ترتيب. في حين تخضع فئات المتغيرات الترتيبية إلى ترتيب. كما تناولنا في هذا الفصل العملية التي على ضوئها يتم تحديد ما إذا كان هذا المتغير متغيراً متصلاً (العمر) أو متغيراً منفصلاً (الدخل).

أسئلة للمراجعة:

1- بين مستوى القياس للمتغيرات التالية:

أ- العرق:

ليبي آسيوي

أمريكي لاتيني

أخرى تذكر

ب- الأمانة:

في طريقك إلى مدرجات الجامعة لاحظت محفظة واقعة على الأرض، وتحتوي هذه المحفظة على بعض النقود والبطاقات الشخصية: هذا الموضوع يمكن تصنيفه إلى الفئات التالية:

- إرجاع المحفظة بكل محتوياتها.
- إرجاع المحفظة ولكن الاحتفاظ بالنقود.
- لا يمكن إرجاع المحفظة على الإطلاق.

2- لقد تم طرح مجموعة من الأسئلة على مجموعة من الناس حول:

- عدد السنوات التعليمية التي تم إنجازها.
- عدد الأطفال.
- عدد الأطباء لكل 1000 / من السكان.

3- هذه مجموعة من المتغيرات تم تصنيفها في فئات:

بين ما إذا كانت متغيرات مقولية أم عددية، مقرراً بعد ذلك ما إذا كانت هذه المتغيرات اسمية أم ترتيبية؟

أ- في العادة، كم مرة تزور أقاربك؟

- أكثر من مرة في الأسبوع.
- مرة في الأسبوع.
- مرة أو مرتين في الشهر.

- مرة في السنة.
- لا أزورهم على الإطلاق.

ب- هل أنت:

عازب - متزوج - مطلق - أرمل؟

ج- ما هو المستوى الدراسي الذي تم إنجازه؟

الابتدائية - الإعدادية - الثانوية - الجامعة فما فوق - لا شيء.

د- هل توافق على تولي المرأة لموقع قيادي (رئيس وزراء مثلاً)؟

- أوافق بشدة.
- أوافق.
- الحياد.
- لا أوافق بشدة.
- أوافق.

هـ- كم دخلك الشهري؟

و- كم عدد أطفالك؟

4- بين ما إذا كانت المتغيرات التالية متغيرات منفصلة أو متصلة:

- كم عدد الأفراد الذين يعيشون في هذا المنزل؟
- كم عدد السنوات التي قضيتها في هذه المهنة؟
- هل توافق أو لا توافق على توزيع الثروة على الليبيين؟
  - أوافق بشدة.
  - أوافق بعض الشيء.
  - لا أوافق.
  - لا أوافق بشدة.

- هل تصنف نفسك حضري أم ريفي؟
- ما هو دخلك في العام الماضي؟

5- أربعة مقاييس تم تقديمها في هذا الفصل: الاسمي / الترتيبي، ذو المسافات والنسبي.

- ما هي المعلومات الإضافية التي نتحصل عليها من خلال المقياس الترتيبي مقارنة بالمقياس الاسمي؟
- ما هي المعلومات الإضافية التي نتحصل عليها من المقياس ذي المسافات إذا ما قورنَ بالمقياس الترتيبي؟
- ما هي المعلومات الإضافية التي نتحصل عليها من المقياس النسبي إذا ما قورنَ بالمقياس ذي المسافات؟

6- أجريت تجربة لمعرفة تأثير عملية القلق على التذكر. وقسمت التجربة إلى مجموعتين: مجموعة يغلب عليها القلق. والمجموعة الثانية غير ذلك أي ليست قلقة. المشاركون بشكل ثابت يتذكرون الفقرات القلقة أكثر من أولئك المشاركين غير القلقين.. بين:

- المتغير المستقل في هذه الدراسة؟
- ما هو مستوى القياس لهذا المتغير المستقل؟
- بين المتغير التابع؟
- ما هو المقياس المستخدم لقياس المتغير التابع؟

الهوامش والمصادر:

أولا الهوامش:

1- عبد الله عامر الهمالي، أسلوب البحث الاجتماعي وتقنياته، منشورات جامعة قاريونس، ط 3، 2003، ص 74.

2- يمكن الرجوع إلى عبد الله عامر الهمالي، أسلوب البحث الاجتماعي وتقنياته، المرجع السابق نفسه، في الجزء المتعلق بالتعريفات التصورية والتعريفات الإجرائية، ص ص 35 - 41.

3- George Argyrous, Statistics for Social and Health Research, With Guide to SPSS, Sage Publications, London, PP. 1 - 9.

4- Ibid, P. 7.

5- Ibid, PP. 7 - 9.

6- J. Richard Kendrick, Social Statistics: An Introduction using SPSS, Second ed, USA, 2005, PP. 35 - 48.

and George Argyrous, Statistics for Social and Health Research, op. Cit., PP. 9 - 14.

ثانيا: المصادر:

1- George Argyrous, Statistics for Social and Health Research With Guide to SPSS, Sage Publications, London, 2001 .

2- J. Richard Kendrick, Social Statistics: An Introduction using SPSS, Second ed, USA, 2005.

3- عبد الله عامر الهمالي، أسلوب البحث الاجتماعي وتقنياته، منشورات جامعة قاريونس، 2003م.

4- ـــــــــــــــ، التقنيات الإحصائية ومناهج البحث، منشورات جامعة قاريونس، 2008م.

الفصل الثالث

تحليل البيانات التكرارية

مقدمة

نركز في هذا الفصل على الإجراءات المستخدمة في تلخيص المعلومات الهائلة التي يتحصل عليها الباحث من خلال إجراء بحثه. ويطلق على هذه الإجراءات، الإحصاءات الوصفية. وتشير الإحصاءات الوصفية إلى الأعداد والتمثيل البياني، وتقنيات الجدولة لتنظيم البيانات وعرضها وتحليلها.

ومن المميزات الكبرى للإحصاءات الوصفية أنها تلخص ذلك الكم الهائل من البيانات حتى يسهل قراءتها بشكل يَسِير. وباختزال كمية البيانات الكبيرة إلى إحصاءات قليلة أو في شكل تمثيل بياني أو جداول تكرارية؛ فإن نتائج البحث سَتُعْرَضُ بشكل مختصر ودقيق.

نفترض أننا قمنا بإجراء مسح اجتماعي لغرض الحصول على بيانات حول دخل عشرين أسرة. وجاءت نتائج هذا المسح كالتالي [1]:

520	462	400	360	0
560	470	420	375	0
700	475	425	400	250
1020	562	450	400	300

إن هذا الترتيب لقياس الدخل يطلق عليه توزيع، وقد تم تقديم هذه البيانات في شكل بيانات خام Raw data كنتائج لهذا المسح. لاحظ أنه إذا عُرِضَت البيانات في شكل بيانات خام، فإن ذلك لا يعطينا أي معنى دال للمتغير الذي نسعى لدراسته. فالبيانات حول دخل هذه الأسرة وبالطريقة التي عُرِضَتْ بها هذه الدخول لهذه الأسر لا معنى لها.

ثمة طريقة أخرى يمكننا اعتمادها في عرض البيانات. وهذه الطريقة يمكن من خلالها عرض هذه البيانات في شكل يمكن معه توليد أرقام قليلة - إحصاء - بحيث تكون هذه المعلومات ذات صلة بما تحتويها البيانات الخام. فالإحصاءات الوصفية عادة ما تمدنا بصورة مهمة للتوزيع الذي لم يكن واضحاً عندما عُرِضَت البيانات في شكلها الخام.

إن أحد هذه الصور الواضحة التي سنتناولها في الفصول اللاحقة من هذا الكتاب ترتبط بفكرة المعدل. فعلى سبيل المثال، يمكننا حساب رقم مفرد لمتوسط الدخل، ويُقَدم هذا الرقم المفرد كجزء من نتائج البحث.

إن مقياس المتوسط الذي تم اختياره لا يصور لنا كل البيانات التي تحتويها البيانات الأولية، وإنما يعطي هذا المتوسط فكرةً عامةً على ما تكون عليه هذه الحالات العشرون، مع السماح للباحث بإيجاد بعض التفسيرات الدّالة.

أنواع الإحصاءات الوصفية:

ما أشرنا إليه حتى الآن هو تقديمنا لفكرة المعدل كصورة مهمة لتوزيع الدرجات التي يرغب الباحث في معرفتها، وبمصطلحات تقنية يطلق علي هذا الأمر مقاييس النزعة

المركزية، ومقاييس التشتت كإحصاءات وصفية. ويطلق على هذين النوعين من الإحصاءات الوصفية: التقنيات العددية Numerical Technigues لوصف البيانات لأنها تتضمن استخدام المعادلات الرياضية لإجراء العمليات الحسابية من البيانات الخام (سنتناول هذه التقنيات العددية في الفصول اللاحقة من هذا الكتاب).

إضافة إلى هذه التقنيات، توجد جملة من الطرق الأخرى التي يمكن من خلالها عرض البيانات ليسهل قراءتها. ومن هذه الطرق المستخدمة في عرض البيانات طريقة بناء الجداول التي تصف توزيع الحالات عبر مدى من القيم لمتغير ما. وكذلك الإحصاءات الثنائية الوصفية مثل: الجداول الثنائية Bivariate Tables، ورسم شكل الانتشار Scatter Plots، ومقاييس التطابق Measures of association. ويمكن تلخيص هذه الطرق المتعددة لوصف البيانات في الجدول التالي:

جدول (3-1): أنماط الإحصاء الوصفي

النمط Type	الوظيفة Function
• الجداول	تمدنا بتوزيع تكراري لمتغير
• التمثيل البياني	يمدنا بعرض بصري لتوزيع متغير ما
• مقاييس النزعة المركزية	حساب معدل الدرجة لتوزيع
• مقاييس التشتت	حساب تباين الدرجات لتوزيع
• مقاييس التطابق	تشير إلى وجود واتجاه وقوة العلاقة بين متغيرين أو أكثر

المصدر: George Argyrous, Statistics for Social and Healh Research, op. Cit, P. 40.

آخذين في الاعتبار أن هناك عدداً كبيراً من الإحصاءات الوصفية المتوفرة. والسؤال الذي يمكن طرحه هنا هو كيف يمكن للباحث أن يقرر ما هي الإحصاءات الوصفية التي يمكن استخدامها في سياق بحث محدد؟ قد يلجأ الباحث في بعض الأحيان إلى تلخيص بياناته من خلال استخدام الجداول. وفي أحيان أخرى إلى استخدام المعدلات و/أو

مقاييس التشتت. بمعنى آخر، إن اختيار الإحصاءات الوصفية لتلخيص البيانات تعتمد بالدرجة الأولى على السؤال البحثي المطروح. وبالرغم من هذا كله، هناك بعض العوامل المحددة تستخدم كطرق إرشادية تساعد الباحث في اختيار الإحصاءات الوصفية. فعلى سبيل المثال، يعتبر مستوى القياس ذا أهمية بالغة في تحديد نوعية الإحصاءات الوصفية المستخدمة الذي سنتناوله بالتفصيل في موضعه في الفصول اللاحقة.

الجداول التكرارية:

الجداول التكرارية هي تلك الجداول التي تلخص إجابات المبحوثين في كل فئة من فئات المتغير المدروس. وببساطة هي عبارة عن قوائم لتوزيع المبحوثين (عدد الإجابات) في كل فئة من فئات المتغير. وتعتبر الجداول التكرارية طريقة شائعة الاستخدام في وصف البيانات للحصول على نتائج بحثية دالّة. أي أنها جداول تلخص توزيع متغير ما وذلك بحصر الحالات في كل فئة من فئات هذا المتغير. كما تساعد هذه الجداول أيضاً في تنظيم البيانات وتحليلها، وبالتالي تمثل الجداول التكرارية الخطوة الأولى في أي تحليل إحصائي. ومن خصائص التوزيع التكراري أنه يمكن استخدامه لكل أنواع المتغيرات، سواء كانت متغيرات عددية أو متغيرات مقولية، متصلة أم منفصلة.

ويستخدم الباحثون الجداول التكرارية لتلخيص الخصائص الديموغرافية مثل: العمر، النوع، الخلفية الحضارية، الحالة الاجتماعية... الخ. وذلك عندما يود الباحث استخدام هذه المتغيرات الديموغرافية لتحليل الاتجاهات والمواقف لأولئك المبحوثين الذين يرغب الباحث في دراستهم. كما تساعد الجداول التكرارية على معرفة الأنماط العامة للبيانات. كما تساعد التوزيعات التكرارية في تقييم تشتت الإجابات للمتغير ليقرر الباحث ما إذا كانت هذه التوزيعات توزيعات متغايرة أم توزيعات متجانسة. فالتوزيعات المتغايرة هي تلك التوزيعات التي يكون فيها توزيع المبحوثين حول المتغير يكاد يكون توزيعاً متساوياً، مشتتاً، أو منتشراً عبر كل فئة من فئات المتغير. ففي التوزيع المتغاير الكامل تكون فيه كل فئة من فئات المتغير لديها نفس (نسبة) عدد المبحوثين. فعلى سبيل المثال، متغير النوع تكون توزيعاته متغايرة 50 % من الذكور، و 50% من الإناث.

تجدر الإشارة إلى أنه كلما زاد انحراف التوزيع أكثر من تساوي تشتت المبحوثين عبر فئات المتغير، قلت عملية التغاير.

وعلى الجانب الآخر، فالتوزيعات المتجانسة هي تلك التوزيعات التي يتركز فيها المبحوثون في فئات قليلة فقط للمتغير. أي أن توزيع التجانس الكامل يعني أن كل المبحوثين يتمركزون في فئة واحدة للمتغير مثل: النوع: ذكور (100) والإناث (0)، الوضع الاجتماعي: منخفض (0)، طبقة عاملة (100)، وسطى (0)، عليا (0).

أشكال التوزيعات التكرارية(*) :

تتخذ التوزيعات التكرارية عدة أشكال منها:

1- جداول البيانات المدرجة Listed data tables
2- جداول التوزيعات التكرارية البسيطة Simple frequency tables
3- جداول التوزيعات التكرارية النسبية Relative frequency tables
4- جداول التوزيعات التكرارية المتجمعة Cumulative frequency tables

1- جداول البيانات المدرجة:

ولتوضيح جداول البيانات المدرجة نفترض أننا تحصلنا على النتائج التالية من خلال مسح افتراضي لعدد 20 حالة. وأن الباحث يرغب في التركيز على ثلاثة متغيرات هي النوع (تم قياسه على المستوى الأسمي)، والرضا المهني (ترتيبي)، وأخيراً متغير العمر (ذو المسافات المتساوية والنسبي). ويوضح الجدول (3 – 2) هذه البيانات الافتراضية.

(*) يمكن للباحث أن يبني هذه الجداول يدوياً إذا كان عدد الحالات التي يتعامل معها صغيرة أو باستخدام برنامج Spss.

جدول (3-2): نتائج المسح الافتراضي لعدد 20 حالة

رقم الحالة	النوع	الرضا المهني *	العمر
1	ذكر	متمتع بصحة جيدة جداً (3)	18
2	ذكر	متمتع بصحة جيدة جداً (3)	21
3	أنثى	تتمتع بصحة جيدة (2)	20
4	ذكر	غير متمتع بصحة جيدة (1)	18
5	أنثى	متمتعة بصحة جيدة جداً (3)	19
6	ذكر	غير متمتع بصحة جيدة (1)	18
7	أنثى	متمتعة بصحة جيدة جداً (3)	22
8	ذكر	متمتع بصحة جيدة جداً (3)	19
9	أنثى	متمتعة بصحة جيدة جداً (3)	18
10	ذكر	متمتع بصحة جيدة جداً (3)	20
11	ذكر	غير متمتع بصحة جيدة (1)	18
12	أنثى	متمتعة بصحة جيدة جداً (3)	19
13	ذكر	متمتع بصحة جيدة جداً (3)	22
14	ذكر	متمتع بصحة جيدة جداً (3)	19
15	أنثى	غير متمتعة بصحة جيدة (1)	20
16	أنثى	متمتعة بصحة جيدة (2)	18
17	ذكر	غير متمتع بصحة جيدة (1)	21
18	أنثى	غير متمتعة بصحة جيدة (1)	19
19	ذكر	متمتع بصحة جيدة (2)	18
20	ذكر	متمتع بصحة جيدة جداً (3)	20

* راضٍ جداً (3)، راضٍ (2)، غير راضٍ (1).

المصدر: George Argyrous, op.cit,P.41.

يطلق على مثل هذا الجدول، جدول البيانات المدرجة في قائمة، باعتبار أن كل درجة لكل حالة لكل متغير تم إدراجه منفصلاً. أي كل حالة في فئة واحدة فقط دون غيرها.

إن مثل هذا النوع من الجداول به عدد كبير من الصفوف بقدر الحالات، وعدد كبير من الأعمدة بقدر المتغيرات التي يتم مشاهدتها. إن مثل هذا الشكل من عرض البيانات لا يكشف لنا عن معلومات مفيدة، خاصة عندما يكون لدينا عدد كبير من الحالات يصبح التعامل مع هذا العدد من الحالات الكبيرة غير عملي. فقد يكون في مقدور الباحث أن يتعامل مع عدد محدود من الحالات، ولكن الأمر لن يصبح ميسوراً إذا زاد عدد الحالات عن عشرين. فإذا تصورنا أنه لدينا مسح لعدد 2000 شخص بدلاً من عشرين، فإنه ليس في إمكاننا بناء جدول بيانات مُدرجة (المساحة). إن الميزة المتعلقة بالبيانات الخام لكل حالة منفصلة، إنها تتيح للباحث إمكانية إجراء عدة حسابات متعلقة بالإحصاء الوصفي.

2- جداول التوزيعات التكرارية البسيطة:

تُسْتَخْدَمُ جداولُ التوزيعات التكرارية البسيطة لوصف البيانات المتعلقة بقيم متغير ما تحت الدراسة والمقاس على المستوى الاسمي والترتيبي وذي المسافات والنسبي. ومن البيانات الخام التي تم تقديمها أعلاه يمكننا بناء مجموعة من الجداول المنفصلة لكل متغير من المتغيرات الثلاثة.

جدول (3-3): جنس المبحوثين

النوع (X)	التكرار (F)
ذكور	12
إناث	8
المجموع	20

جدول (3ـ4): الرضا الصحي للمبحوثين

الرضا الصحي (X)	التكرار (F)
غير راضٍ	7
راضٍ	5
راضٍ جداً	8
المجموع	20

جدول (3ـ5): عمر المبحوثين

العمر بالسنوات (X)	التكرار (F)
18	7
19	5
20	4
21	2
22	2
المجموع	20

تشير هذه الجداول إلى مجرد البناء الأدنى الذي ينبغي أن تحتويها كل الجداول التكرارية بغض النظر عن المتغير ومستوى قياسه. ويتمثل الحد الأدنى لبناء الجداول التكرارية البسيطة في:

- أن يكون للجدول عنوان يحدد بوضوح أسماء الفئات.
- حصر الوقائع المتعلقة بكل فئة من فئات المتغير موضوع الدراسة بحيث تكون شاملة ومانعة التبادل.
- الحصر الكلي للحالات (N).

ويلاحظ في الجدول المتعلق بتوزيع المبحوثين حسب النوع أن الصف الأول يحتوي على الذكور، في حين يحتوي الصف الثاني على الإناث. وهذا التصنيف تصنيف اعتباطي يتم استخدامه في حالة المقياس الاسمي. ويمكننا أن نعيد تنظيم هذه الفئات بأي طريقة أخرى يختارها الباحث. ولقد تم وضع الذكور أولاً باعتبار أن ذلك ممارسة عامة في المقياس الاسمي على أن تنظم الصفوف بكيفية تكون الفئة ذات التوزيع الأعلى في الصف الأول. والفئة التي تليها من حيث العدد في الصف الثاني وهكذا... ويرجع السبب في هذا الترتيب إلى "فئة نموذج" Model Category الذي يستخدم في تحليل توزيعات المتغير الاسمي. في حين أن توزيع المبحوثين حسب درجة الرضا الصحي، والعمر فإن فئاتهما تخضع إلى قاعدة صارمة وفقاً للحقيقة أن التعامل معهما وفقاً للمقياس الترتيبي وذي المسافات والنسبي. حيث إن الإجراء المتبع في هذه المستويات من القياس، عامة، يبدأ بأقل قيمة في التوزيع ثم ترتقي هذه القيم تدريجياً في القياس إلى الأسفل بحيث نتحرك في الجدول صفاً بصف(2).

3- جداول التكرارات النسبية:

تعبر الجداول التكرارية النسبية عن عدد الحالات داخل كل قيمة من قيم المتغير كنسبة مئوية، أو تناسب من المجموع الكلي للحالات. ولكي نولد جدولاً تكرارياً نسبياً من البيانات السابقة التي تعرضنا لها، ينبغي على الباحث أن يلم بتقنيات النسب المئوية والتناسب.

أـ النسب المئوية:

النسب المئوية عبارة عن إحصاءات تقيس العدد الكلي للحالات استناداً لقاعدة قيمتها (100). والمعادلة التالية تُسْتَخْدَمُ لحساب النسبة المئوية:

$$\% = \frac{F}{N} \times 100$$

حيث إن: F يشير إلى التكرار وعدد الحالات في أي قيمة معينة.

n تشير إلى العدد الكلي للحالات لجميع الفئات.

ولتوضيح العملية التي يتم بها حساب النسب المئوية، نورد المثال التالي:

نفترض أن لدينا إحصاءاتٍ تشير إلى الوضع الاجتماعي لمجموعة من الأفراد يبلغ عددهم 133 فرداً، وكانت أوضاعهم الاجتماعية كالتالي:

عدد المتزوجين 63 حالة، والمطلقين 19 حالة، والمخطوبين 27 حالة، والأرامل 13 حالة، وآخرون 11 حالة. فإن النسب المئوية لهؤلاء تحسب كما يلي [3]:

- المتزوجون = $\frac{63}{133}\times100 = 47.3\,\%$
- المطلقون = $\frac{19}{133}\times100 = 14.3\,\%$
- المخطوبون = $\frac{27}{133}\times100 = 20.3\,\%$
- الأرامل = $\frac{13}{133}\times100 = 9.8\,\%$
- آخرون = $\frac{11}{133}\times100 = 84.3\,\%$

ب- التناسب Proportions:

يوفر التناسب والنسب المئوية إطاراً مرجعياً بطريقة تجعل من البيانات أكثر وضوحاً. إلا أن الشيء الذي يمكن ملاحظته عند التعامل مع التناسب (P)، يستند على أساس قاعدة واحد صحيح، بدلاً من مائة كما هو الحال في النسبة المئوية.

$$P = \frac{F}{N}$$

إن النتيجة التي سنتحصل عليها من خلال التناسب يمكن التعبير عنها من خلال التعبير العشري decimal. ففي المثال السابق:

- المتزوجون = $\frac{63}{133} = 0.473$

- المطلقون = $\frac{19}{133} = 0.143$
- المخطوبون = $\frac{27}{133} = 0.203$
- الأرامل = $\frac{13}{133} = 0,098$
- آخرون = $\frac{11}{133} = 0.083$

وتجدر الإشارة هنا - بشكل عام - إلى أن التعامل مع النسب المئوية أسهل منه مقارنة بالتعامل مع التناسب؛ ولسبب أو آخر، فإن الناس يفضلون التعامل مع الرقم الكلي بدلاً من العشريات. إلا أننا وفي ثنايا هذا الكتاب سنتعامل بشكل كبير مع التناسب ومن هنا ينبغي على القارئ أن يتعلم العلاقة البسيطة بين التناسب والنسب المئوية المألوفة [4].

ولتحويل التناسب إلى ما يقابله من قيمة مئوية يكون بتحريك النقطة العشرية مكانين إلى اليمين (هذه العملية تكون تماماً لو ضربنا الرقم في مائة). ولكي تُحَوَّلَ النسبة المئوية إلى ما يقابلها من تناسب يتم ذلك بتحريك النقطة العشرية إلى اليسار (هذه العملية تعطي نفس النتيجة إذا تم تقسيم النسبة المئوية على مائة).

ثمة بعض المحاذير ينبغي على الباحث أن يضعها في الاعتبار عند التعامل مع التناسب والنسب المئوية. أولهما عليه النظر عندما يواجه النسب أو التناسب هو مجموع الصف الذي تم حساب النسب والتناسب منه. ذلك أن النسب المئوية والتناسب تستخدم في بعض الأحيان لإخفاء فروق شديدة في الحجم المطلق. فالزيادة على سبيل المثال، في معدلات البطالة من 10 % إلى 10.5 % لا تبدو زيادة شديدة بالمفهوم الإحصائي، في حين أن نسبة 0.5 تمثل 35.000 شخصاً، وبالتالي تعتبر هذه الزيادة بمفهوم الوضع الاقتصادي والاجتماعي زيادة كبيرة.

وبشكل معاكس، فإن التغير الكبير في أرقام النسب المئوية يمكن أن يكون أمراً عادياً

عند التعامل بأرقام مطلقة. وعلى سبيل المثال: فإن عدد الناس الذين حضروا الاجتماع الأخير يزيد بنسبة 150 في المائة أكبر من نسبة أولئك الذين حضروا الاجتماع السابق ولكن إذا كان هذا فعلياً يرجع إلى حضور خمسة أفراد في الاجتماع الأخير بدلاً من شخصين في الاجتماع السابق، فإنه قليلاً ما يكون ازدياداً شديداً. فعند التعامل مع أرقام مطلقة صغيرة، فإن إضافات صغيرة إما للمجموع أو للفئات المكونة لهذا المجموع سوف تؤثر كثيراً في الرقم النسبي المحسوب[5].

وعليه، عند التعامل مع عدد صغير من الحالات أي أقل من عشرين، ففي هذه الحالة يفضل الاعتماد على التكرارات الفعلية بدلاً من النسب المئوية والتناسب [6].

بعد أن تعرفنا على النسب المئوية والتناسب واستخدامها في بناء جداول التوزيعات التكرارية النسبية للبيانات التي تم استخدامها سابقاً. يمكننا أن نضيف للجدول لكل متغير، عموداً يوضح النسبة المئوية (أو التناسب) للحالات التي تقع في كل فئة.

الجدولان التاليان يوضحان العملية الحسابية المتعلقة بتوليد التوزيعات النسبية.

جدول (3-6): نوع المبحوثين

النوع (X)	التكرار (F)	النسب المئوية %
أنثى	8	$\frac{8}{20}\times 100 = 40$
ذكر	12	$\frac{12}{20}\times 100 = 60$
المجموع	**20**	**% 100**

جدول (3ـ7): الرضا الصحي

الرضا الصحي (X)	التكرار (F)	النسب المئوية %
غير راضٍ	7	$\frac{7}{20} \times 100 = 35$
راضٍ	5	$\frac{5}{20} \times 100 = 25$
راضٍ جداً	8	$\frac{8}{20} \times 100 = 40$
المجموع	**20**	**% 100**

جدول (3ـ8): عمر المبحوثين

العمر بالسنوات (X)	التكرار (F)	النسب المئوية %
18	7	$\frac{7}{20} \times 100 = 35$
19	5	$\frac{5}{20} \times 100 = 25$
20	4	$\frac{4}{20} \times 100 = 20$
21	2	$\frac{2}{20} \times 100 = 10$
22	2	$\frac{2}{20} \times 100 = 10$
المجموع	**20**	**% 100**

تجدر الملاحظة أن عمود النسب المئوية لابد أن يصل مجموعه إلى 100 %، لكننا في بعض الأحيان نجد أن النسبة في الجدول قد لا يصل مجموعها إلى 100 % وذلك نتيجة لعملية التقريب. فعلى سبيل المثال، قد تكون النسبة 22.3 %، 38.4 %، 39.3 %. فإذا تم تقريب هذه الأرقام إلى أقرب وحدة عشرية: 22 %، 38 %، 39 %، فإن هذه النسب المقربة يصل مجموعها إلى 99 %. وعندما يحدث ذلك، ينبغي على الباحث أن يبين في هامش تحت الجدول مشيراً بعبارة "النسبة لم تصل إلى 100، وذلك نتيجة لعملية التقريب" [7].

4- جداول التكرار المتجمع:

يمكننا القول بصفة عامة إن بناء جداول تكرارية لمتغيرات مقاسة على المستوى الترتيبي، وذي المسافات والنسبي إذا ما قورنت ببناء الجداول التكرارية البسيطة. ففي الجدول التكراري المتجمع تضاف أعمدة تتعلق بالتكرارات المتجمعة، والتكرارات المتجمعة النسبية. ولما كانت البيانات المقاسة على المستويين الترتيبي وذي المسافات والنسبي تسمح لنا بعملية ترتيب الحالات المدروسة من الأدنى إلى الأعلى، فإنه يفضل في هذه الحالة معرفة العدد و/أو النسبة للحالات التي تقع فوق أو تحت نقطة محددة من المقياس. فالجدول التكراري المتجمع يشير إلى كل قيمة في التوزيع وتجميع أعداد الفئات إلى بعضها البعض لتشمل تلك القيمة.

أما الجدول التكراري المتجمع النسبي Cumulative relative frequency table فهو يبين كل فئة في التوزيع، النسب المئوية والتناسب للعدد الكلي للحالات تجمع بعضها إلى بعض لتشمل تلك القيمة.

تجدر الملاحظة أن كل التكرارات المطلقة والنسبية والتكرارات المتجمعة لمتغير ما يمكن جمعها – في بعض الأحيان – في جدول واحد كما هو بيّن في الجدولين التاليين:

جدول (3ـ9) الرضا الصحي للمبحوثين

معدل الرضا الصحي	التكرار	التكرار المتجمع	النسبة المئوية %	النسبة المتجمعة %
غير راضٍ	7	7	35	$\frac{7}{20}\times100=35$
راضٍ	5	12	25	$\frac{5}{20}\times100=25$
راضٍ جداً	8	20	40	$\frac{8}{20}\times100=40$
المجموع	**20**		**100**	**% 100**

جدول (3 ـ 10) عمر المبحوثين بالسنوات

العمر بالسنوات	التكرار	التكرار المتجمع	النسبة المئوية %	النسبة المتجمعة %
18	7	7	35	35
19	5	12	25	60
20	4	16	20	80
21	2	18	10	90
22	2	20	10	90
المجموع	**20**		**100**	**% 100**

وبتلخيص هذه التوزيعات بهذه الطريقة يمكن للباحث أن يجيب بشكل محدد على الأسئلة البحثية المطروحة. فإذا رغب الباحث في معرفة كم عدد المبحوثين الذين يبلغون من العمر 18 أو 19 سنة.

فالباحث ببساطة ينظر إلى مجموع الحالات في أول الصفين من الجدول رقم (10) ليجد أن التكرار المتجمع يصل إلى 12، أي 60% من كل الحالات. وبالطريقة نفسها، إذا رغب الباحث في معرفة كم عدد الحالات التي تزيد أعمارها عن 19 سنة، فإنه باستطاعته ملاحظة أن 60% من الحالات أعمارهم 19 سنة أو أقل، فإن ذلك يترك 40% من الحالات فوق هذا العمر (40% = 60 − 100)[8].

تجدر الإشارة إلى نقطة إضافية في هذا السياق هي أنه عند التعامل مع المتغيرات المقاسة على المستوى ذي المسافات والنسبي عادة ما تستخدم فئات متساوية Class Interval بدلاً من قيم فردية لبناء توزيع تكراري.

وعند بناء التوزيع التكراري للفئات المتساوية يتطلب الأمر اختزال البيانات إلى عدد من الفئات الموجزة بدرجة يمكن التعامل معها، إلا أن هذا الاختزال قد يؤدي إلى ضياع بعض المعلومات، وعليه ينبغي على الباحث تجنب اختزال البيانات إذا كانت البيانات أصلاً قليلة وعرضت بشكل بسيط وبقيم فردية.

وَيُسْتَخْدَم التوزيع التكراري للفئات المتساوية إذا كان مدى القيم كبيراً جداً يصعب معه عرض البيانات وتحليلها. ولتوضيح بناء جدول تكراري يحتوي على فئات متساوية نسوق المثال التالي للتدليل:

جدول (3ـ 11)

يحتوي على فئات متساوية: الدخل الأسبوعي لعشرين مبحوثاً

الفئات (X)	التكرار (F)
0 – 99	2
100 – 199	0
200 – 299	1
300 – 399	3
400 – 499	9
500 – 599	3
600 +	2
المجموع	**20**

من هذا الجدول يمكن ملاحظة أن أعلى توزيع للحالات التي تقع داخل الفئة (499 – 400)، كما يلاحظ أيضاً توزيع الدرجات عبر الفئات المتساوية الأخرى. لاحظ أننا في هذا الجدول لن نستخدم القيم الفردية التي تظهر في التوزيع لتبين كل صف، وبدلاً من ذلك تم استخدام حدود فئة محددة. فالحدود الحقيقية هي الحدود العليا والدنيا للفئات المتساوية التي تحدد نطاقاتها. وبشكل عام، إن الفئات المتساوية لديها نفس النطاق (الحجم)، بالرغم من أنه عند نهاية الحد الأدنى والحد الأعلى لمدى البيانات وضعت بشكل غير محدد open - ended كما هو الحال في الفئة من 600 فأكثر. إِن التطابق الفعلي للفئات المتساوية يعتمد على موقف معين، خصوصاً كمية المعلومات المطلوبة. وعليه، فإنه كلما اتسعت الفئات المتساوية سهل قراءة التوزيع، غير أن أقل المعلومات تكون في المتناول، فعلى سبيل المثال، إذا

استخدمنا فئات متساوية، ولنقل بمدى 200 مثل (199 – 0) و (399 – 200... الخ). فإن هذه الفئات ستؤدي في المحصلة النهائية إلى ضياع معلومات كثيرة. فالحالات التي تختلف كثيراً فيما يتعلق بالمتغير المرغوب في دراسته، فالشخص الذي يكسب 400 د.ل، والشخص الذي يكسب 560 د.ل شهرياً يمكن النظر إليهما بشكل متساوٍ. وبشكل عام، عندما نقوم بجمع قيم في فئات متساوية فإننا نفقد معلومات حول التباين الذي تحتويه هذه البيانات، وبالتالي كلما كان نطاق المسافة واسعاً بين الفئات زاد معه بشكل كبير ضياع المعلومات.

وبشكل آخر، إذا كان لدينا حجمُ ضيِّقُ للفئات المتساوية في جدول، فإنه باستطاعتنا أن نكتشف تباين أكبر في البيانات التي نتعامل معها، ولكننا لا نستطيع أن نعرض البيانات بشكل بسيط يجعل منها طيِّعة ومقروءة. فعلى سبيل المثال، إذا استخدمنا فئات متساوية بطول فئة 50 لبيانات الدخل (99 – 50، 49 – 0 إلى آخره) فإن عدد الصفوف في الجدول لن تختزل البيانات في شكل يمكن قراءته بحسب ما نريد. وعند بناء جدول يحتوي على فئات متساوية، يتطلب منا أن نضمن أن تكون المسافات مانعة التبادل Matually Exclusive، وبالتالي فإن اختيارنا لِـ 100 كطول للفئة في الجدول رقم (11) حيث الفئات كالتالي: 99 – 0، 199 – 100، 299 – 200... الخ. لاحظ أن الحد الأعلى لكل فئة لا يتلامس مع الحد الأدنى للفئة التي تليها. ويتضح أن ثمة تفاوت بين 99 و 100، 199 و 200، 299 و 300، وهكذا...

من المفاهيم الأخرى التي يمكن التعامل معها في إطار الفئات المتساوية، هو مفهوم مركز الفئة Mid. Point (M) للفئات المتساوية. ونعني بمركز الفئة، ببساطة، مجموع الحدود الدنيا والحدود العليا مقسماً على 2.

$$Mid.Point\ (m) = \frac{\text{الحد الأعلى + الحد الأدنى}}{2}$$

فمركز الفئة للفئة 99 – 0 هو:

$$M = \frac{0+99}{2} = 49.5$$

والجدول التالي يبين الحدود الحقيقية لدخل عشرين مبحوثا:

الدخل	مركز الفئة (m)	التكرار
0 – 99	49.50	2
100 – 199	149.50	0
200 – 299	249.50	1
300 – 399	349.50	3
400 – 499	449.50	9
500 – 599	549.50	3
600 +	649.50	2
المجموع		**20**

إن بناء جدول يحتوي الحدود الحقيقية ومركز الفئة للفئات المتساوية سيكون مفيداً عند الحديث عن مقاييس النزعة المركزية التي سنتناوله في موضعه من هذا الكتاب.

من التقنيات الأخرى الشائعة في تبويب البيانات في شكل مجموعات يمكن قراءته هي العشيرات Deciles. وبدلاً من استخدام قيم المتغير لتبويب الفئات، فالعشيرات تستخدم نسب محددة لبناء جدول حوله. فمجموع الحالات ينبغي أن يرتب ويُقسم إلى 10 مجموعات متساوية الحجم. إن هذا الإجراء، إجراء شائع الاستخدام في تحليل البيانات المتعلقة بالدخل. فعلى سبيل المثال، يمكننا ترتيب كل العائلات في مجموع محدد طبقاً للدخل، من الأسر الأكثر فقراً إلى الأسر الأغنى. وبعد ذلك تقسم إلى 10 مجموعات متساوية على أن يحتوي العشير الأول، 10 % من العائلات الأكثر فقراً، ويتضمن العشير الثاني 10 % من الأسر الأخرى حتى نصل إلى العشير العاشر الذي يحتوي هو الآخر على 10 % من الأسر الأغنى. وبالنظر إلى النسبة الكلية للدخل التي يحتفظ بها كل عشير، فإنه بإمكاننا أن ندرك توزيع الدخل وطبيعة التغيرات التي حدثت [9].

جدول (3-12) الدخل الإجمالي وفقا للعشير (أستراليا 1989)

العشير	حصة الدخل الإجمالي 1989 %
الأول	1.7
الثاني	2.8
الثالث	3.9
الرابع	5.2
الخامس	6.8
السادس	8.6
السابع	10.7
الثامن	13.4
التاسع	17.3
الأعلى	29.4

المصدر: George Argyrous, op.cit, P.53

من خلال هذه البيانات الواردة أعلاه، يتضح لنا أن توزيع الدخل الإجمالي لم يؤثر بشكل متساوٍ عبر الأسر (طبقاً لهذا المقياس).

إجراءات توليد التكرارات باستخدام SPSS:

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Analyze ⇦ *Descriptive Statistics* ⇦ *Frequencies*

2- اختر المتغير أو المتغيرات لتوليد جدول تكراري بالنقر على هذه المتغيرات: (تقود هذه العملية إلى الحصول على جداول التكرار بشكل تلقائي من خلال لصق أكثر من متغير في صندوق (Variable (s. فإذا أراد الباحث كل المتغيرات (في هذا المثال ثلاثة متغيرات) فعليه لصقها معاً.

3- انقر على ◄ بحيث يتم لصق المتغير أو المتغيرات المختارة في المساحة تحت المتغير أو المتغيرات (Variable (s وهي قائمة المتغيرات التي من خلالها يتولد جدول التكرار.

4- انقر على ok.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

Frequencies

Statistics

	Age in years	Health rating	Sex of respondent
N valid	20	20	20
missing	0	0	0

Frequencies Table

	Frequencies	Percent	Valid percent	Cumulative Percent
Valid 18	7	35.0	35.0	35.0
19	5	25.0	25.0	60.0
20	4	20.0	20.0	80.0
21	2	10.0	10.0	90.0
22	2	10.0	10.0	100.0
Total	20	100.0	100.0	

Health rating

Valid percent	**Cumulative Percent**	**Percent**	**Frequencies**	
35.0	35.0	35.0	7	Valid unhealthy
60.0	25.0	25.0	5	healthy
100.0	40.0	40.0	9	Very healthy
	100.0	100.0	20	Total

Sex of respondent

	Frequencies	**Percent**	**Valid percent**	**Cumulative Percent**
Valid Female	8	40.0	40.0	40.0
Male	12	60.0	60.0	100.0
Total	20	100.0	100.0	

أحياناً يصادف الباحث أن بعض المبحوثين لا يجيبون على بعض الأسئلة ومن هنا تكون لديه بعض المعلومات المفقودة فإذا افترضنا أن الحالة رقم 19 والحالة رقم 20 (ذكور) لم يحددا جنسهم. فالمخرجات في هذه الحالة يمكن بيانها من خلال الإجراء التالي:

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Frequencies ⟸ *Descriptive Statistics* ⟸ *Analyze*

أمر للنوع سوف يولد المخرجات التالية:

Frequencies

Statistics Sex of respondent

N valid Missing	18 2

Sex of respondent

	Frequencies	**Percent**	**Valid percent**	**Cumulative Percent**
Valid Female	8	40	44.4	44.4
Male	10	50.0	55.6	100.0
TOTAL	18	90.0	100.0	
Missing	2			
TOTAL	20			

شكل رقم (3-1) مخرجات Spss للتكرارات

أسئلة للمراجعة:

1- بين الاختلاف بين النسب المئوية والتناسب؟

2- لماذا يكون التناسب دائماً أصغر من قيمة نظيره في النسب المئوية؟

3- حوّل التناسب التالي إلى نسب مئوية:

أ- 0.01 ب- 0.13 ج- 1.24 د- 0.0045

4- حوّل النسب المئوية التالية إلى تناسب:

أ- 12 % ب- 13.4 % ج- 167 % د- 3.5 %

5- من البيانات التالية:

80، 79، 90، 66، 99. 37، 78
81، 53. 78، 71، 57. 68، 89
60، 76، 98، 79. 93، 58. 77
69، 74، 80، 83، 92، 49. 77
75، 48، 73، 74، 84، 62، 90
59، 65، 87، 84، 75، 32، 98
72، 85، 62، 70، 55، 95، 86، 63

المطلوب: بناء جدول يحتوي على فئات متساوية بطول (10)

$$\begin{pmatrix} 30 & - & 39 \\ 40 & - & 49 \end{pmatrix}$$

وهكذا...

الهوامش والمصادر:

أولا: الهوامش:

1- George Argyrous, Statistics for Social & Health Research, with A Guide to Spss, SAGE Publications, London, 2001, P.38.

2- Ibid, P. 43.

3- Hubert M. Blalock. Jr. Social Statistics, Mac Graw Hill Book Company, INC, New york, 1972, P.34.

4- Ibid, P. 46.

5- Ibid, P. 46.

6- مصطفى خلف عبد الجواد، الإحصاء الاجتماعي: المبادئ والتطبيقات، دار المسيرة للنشر والتوزيع والطباعة، عمان، 2009، ص 40.

7- George Argyrous, Statistics for Social & Health Research, op. Cit, P.47.

8- Ibid, PP. 47 - 48.

9- Ibid, P. 51, 52 ,53.

ثانيا: المصادر:

1- George Argyrous, Statistics for Social & Health Research, with A Guide to Spss, SAGE Publications, London, 2001.

2- Hubert M. Blalock Social Statistics, Mac Graw Hill Book Company,INC, New york, 1972 .

3- Joseph F. Healey, The Essentials of Statistics A Tool for Social Research, Wadsworuh Cen Gage Learning, USA. 2010.

4- J. Richard Kendrick, Social Statistics: AN Introdaction using SPSS for Windows 2 ed, USA, New york, 2005.

5- مصطفى خلف عبد الجواد، الإحصاء الاجتماعي: المبادئ والتطبيقات، دار المسيرة للنشر والتوزيع والطباعة، عمان، 2009.

الفصل الرابع

مقاييس النزعة المركزية ومقاييس التشتت

أولاً : مقاييس النزعة المركزية :

تحتوي مقاييس النزعة المركزية على المنوال والوسيط والمتوسط وتساعد هذه المقاييس في إيجاد أنماط من البيانات كما أن لهذه المقاييس استخدامات في الحياة اليومية.

وتشير مقاييس النزعة المركزية إلى قيمة متوسطية في التوزيع وتتخذ هذه المقاييس إشارات مختلفة كما يتبين في الجدول التالي، وأن اختيار هذه المقاييس يمكن أن يتم من خلال حسابه من أي مجموعة من البيانات التي تتغير من خلال المستوى الذي تم على ضوئه قياس المتغير.

جدول (4-1) مقاييس النزعة المركزية

مستوى القياس	مقاييس النزعة المركزية
الاسمي	المنوال
الترتيبي	المنوال + الوسيط
ذو المسافات والنسبي	المنوال + الوسيط + المتوسط

المصدر: George Argyrous, Statistics for Social & Health Research, with a Guide to SPSS, SAGE Publications, London ,2001, p.64

نلاحظ من خلال هذا الجدول أن القواعد الأساسية في الإحصاء: هي تقنيات يمكن تطبيقها على مستوى معين من القياس، وكذلك يمكن تطبيقها على مستوى أعلى من القياس. فعلى سبيل المثال، إن مقياس النزعة المركزية يمكن حسابه من البيانات الاسمية (المنوال) وكذلك يمكن حسابه من البيانات الترتيبية والبيانات ذات المسافات المتساوية والنسبية وبطريقة مشابهة فإن المقاييس التي يمكن حسابها من البيانات الترتيبية يمكن حسابها أيضاً من البيانات ذات المسافات المتساوية والنسبية. وتجدر الإشارة هنا إلى أن المقاييس التي يمكن حسابها من مستوى معين من القياس ليس باستطاعتنا دائماً حسابها من مستويات أقل، فالمتوسط الحسابي على سبيل المثال، يمكن حسابه لمستوى أعلى من القياس (ذي المسافات والنسبي) ولتوضيح ذلك في كيفية حساب كل واحد من مقاييس النزعة يمكننا بيان ذلك من خلال البيانات النظرية التالية:

جدول (4-2) نوع المبحوثين

النوع	التكرار
ذكور	12
إناث	8
المجموع	20

جدول (4-3) الرضا الصحي للمبحوثين

الرضا الصحي	التكرار
غير راض	7
راض	5
راض جداً	8
المجموع	20

جدول (4-4) عمر المبحوثين

عمر المبحوثين بالسنوات	التكرار
18	7
19	5
20	4
21	2
22	2
المجموع	20

من خلال بيانات جدول (4 – 4) يمكننا حساب المنوال والوسيط والمتوسط، في حين يمكننا من الجدول (4 – 3) حساب المنوال والوسيط، أما في جدول (4 – 2) فإنه ليس باستطاعتنا سوى حساب المنوال.

المنوال:

دعنا نبدأ بمقياس المنوال (Mode) وهو أبسط مقاييس النزعة المركزية حيث يمكن حسابه على كل مستويات القياس.

ويعرف المنوال: بأنه القيمة الأعلى تكراراً في التوزيع أي القيم التي تتكرر أكثر من غيرها. والمنوال هو المقياس الوحيد للنزعة المركزية الذي يمكن حسابه من البيانات الاسمية والترتيبية وذات المسافات المتساوية والنسبي، وتعتبر هذه ميزة على الخيارات الأخرى (الوسيط والمتوسط) وذلك لسهولة حسابه. فإذا نظرنا إلى التوزيع التكراري لعشرين حالة حسب النوع (ذكور 12 حالة) و (إناث 8 حالات) فإن أعلى تكرار لهذه البيانات هم فئة الذكور، ومن هنا فإن التوزيع الأعلى تكراراً هو توزيع الذكور (Mo).

تجدر الإشارة إلى أنه عند التعامل مع المنوال يجب أن نضع في أذهاننا نقطة أساسية وهي أن المنوال هو قيمة المتغير الأكثر تكراراً وليس عدد الأوقات التي تظهر فيها القيمة في التوزيع[1].

ثمة خاصية ترتبط بالمنوال قد لا تنطبق على الوسيط أو المتوسط كمقياسين من مقاييس النزعة المركزية باعتبار أن التوزيع المرتبط بالمنوال قد يحتوي على أكثر من منوال، فعلى سبيل المثال، نفترض أن لدينا التوزيع التالي المتعلق بمتغير العمر كما هو موضح في الجدول (5).

جدول (4-5) عمر المبحوثين

العمر بالسنوات	التكرار
18	7
19	5
20	4
21	2
22	7
المجموع	**25**

نلاحظ من خلال هذا الجدول، أن فئتين من فئات العمر تحتويان على أعلى توزيع: 18 سنة و 22 سنة ويطلق على مثل هذا التوزيع التوزيع المزدوج Bimodel أما الوسيط أو المتوسط على الجانب الآخر، فهما دائماً عدداً مفرداً كمتوسط بغض النظر عن التوزيع. وتجدر الإشارة في هذا السياق إلى أنه إذا ما اسْتُخْدِمَ المنوال لوصف مجموعة من البيانات ذات المسافات المتساوية والنسبية فإن هذا الاستخدام يواجه مشكلة أساسية وهي إذا أخذنا على سبيل المثال، الدرجات التالية التي توضح الوقت بالثواني لكي يأخذ الدواء تأثيره على عينة من المرضى ورتبت هذه البيانات وفقاً للترتيب التالي:

98، 97، 95، 91، 89، 86، 84، 82، 81، 36، 36، 33

وبنظرة سريعة لهذه البيانات نجدها تتمركز في مكان ما بين 90 – 80 وأن المنوال لهذا التوزيع هو 36 ثانية حيث ظهر هذا الرقم مرتين في التوزيع، في حين أن الأرقام الأخرى قد ظهر كل منها مرة واحدة في التوزيع وبشكل واضح يتبين لنا أن المنوال لا يعكس حقاً النزعة المركزية لهذا التوزيع وفي مثل هذه الأحوال ينبغي علينا استخدام إما مقاييس النزعة المركزية الأخرى كتلك التي سنناقشها في حينه، أو تنظيم البيانات في شكل فئات متساوية وبيان نموذج هذه الفئات المتساوية بدلاً من نموذج الدرجات الفردية[2].

الوسيط:

الوسيط لمجموعة من الأعداد المرتبة ترتيباً تنازلياً أو تصاعدياً، هو العدد الأوسط منها إذا كانت الأعداد فردية. أما إذا كانت الأعداد زوجية، فالوسيط يقع بين قيمتين أي أنّ الوسيط هو متوسط هاتين القيمتين ويمكن حساب درجة الوسيط بعدة طرق مختلفة اعتماداً على طبيعة البيانات المتوفرة لدى الباحث ولحساب الوسيط يتطلب ترتيب الحالات من الأسفل إلى الأعلى طبقاً للكمية التي يمتلكها المتغير لكل حالة. فإذا رتبت الحالات في التوزيع من الأدنى إلى الأعلى فإن الوسيط هو القيمة التي تقسم البيانات إلى نصفين. أي أنَّ نصف كل هذه الحالات لديها قيمة أكبر من الوسيط ونصف الحالات الأخرى لديها قيمة أقل من الوسيط. بمعنى آخر، إذا ما اخترنا عشوائياً حالة من السلسلة المرتبة ستكون لدينا فرصة 50 %، أن هذه الحالة ستقع فوق الوسيط وأن فرصة 50 % ستقع تحت الوسيط من البيانات التالية يمكننا حساب الوسيط:

12، 64، 56، 3، 87، 25، 93

ولإيجاد الوسيط لهذه البيانات بادئ ذي بدء ينبغي أولاً ترتيب هذه البيانات من الأدنى إلى الأعلى.

93	87	64	56	25	12	3
7	6	5	4	3	2	1

ولما كانت هذه الحالات السبع أرقاماً مفردةً فإن قيمة الوسيط سوف تتحدد بالقيمة الرابعة في هذا التوزيع وهي 56.

$$N+1/2 = 7+1/2 = 4 = 56\ (56)$$

وإذا أضفنا قيمة (98) للتوزيع فإن الترتيب سيتخذ الشكل التالي:

98	93	87	64	56	25	12	3	الدرجة
8	7	6	5	4	3	2	1	الترتيب

في هذا التوزيع سيكون لدينا 8 حالات زوجية فالوسيط لهذه البيانات هو متوسط القيمتين الرابعة والخامسة:

$$N+\frac{1}{2}=\frac{65+64}{2}$$

$$N+\frac{1}{2}=8+1=\frac{9}{2}=4.5=60\left(\frac{65+64}{2}=60\right)$$

أما إذا كانت لدينا بيانات مبوبة أي توزيعات تكرارية مبوبة.

جدول رقم (4-6)

X	M	F	
0 – 2	1	8	20
3 – 5	4	12	
6 – 8	7	6	9
9 – 11	10	2	
12 – 14	13	1	
		N = 29	

في مثل هذه البيانات يمكننا تقدير الوسيط كمساوٍ لمركز الفئة Mid point التي تحتوي على (N+1/2) لهذا النظام الترتيبي ولما كانت N=29 (N+1/2=15) ويمكن إيجاد هذا الرقم في الفئة (5–3) التي تحتوي درجات تقع في موضوع ما بين N9 إلى N20 ولذا فإن الوسيط يقدر بـِ 4 وهي النقطة الوسطى للفئة.

المتوسط:

يشيع استخدام المتوسط الحسابي للبيانات المقاسة على المستوى ذي المسافات والنسبي. ويعرف المتوسط: بأنه مجموع كل الدرجات في التوزيع مقسماً على عدد الحالات. وعند حساب المتوسط الحسابي من المجتمع ككل يشار إليه بـِ u في حين يشار إليه بـِ $\overline{X}$ عند التعامل مع العينات. وأن المعادلة المستخدمة في حسابه تعتمد فيما إذا كانت البيانات قد عرضت في شكل (بيانات خام) أو في شكل (توزيعات تكرارية) أو في شكل (فئات متساوية).

في حالة البيانات الخام التي وضعت في شكل قائمة فإن المعادلة المستخدمة لمتوسط المجتمع ومتوسط العينة ستكون بالشكل التالي:

المجتمع $$u = \frac{\sum X}{N}$$

العينة $$\overline{X} = \frac{\sum X}{N}$$

حيث إن: N تشير إلى حجم المجتمع.

N تشير إلى حجم العينة.

X تشير لكل درجة في التوزيع.

فإذا كانت لدينا البيانات التالية: 21، 19، 15، 12

فالمتوسط الحسابي سيكون:

$$\overline{X} = \frac{\sum X}{N} = \frac{12+15+19+21}{4} = 16.75$$

أما في حالة البيانات التكرارية كما في الجدول (4-7) ففي هذا الجدول نجد أن بيانات العمر لم تعد في شكل بيانات خام وإنما عرضت في شكل توزيعات تكرارية. وفي هذه الحالة يمكننا استخدام المعادلة التالية لحساب متوسط العينة:

جدول (4ـ7)

X	F
18	7
19	5
20	4
21	2
22	2

$$\overline{X} = \frac{\sum FX}{N} = \frac{(18\times 7)+(19\times 5)+(21\times 2)+(22\times 2)}{20}$$

$= 19.35$ سنة

وفي حالة البيانات المبوبة فإن الجدول التكراري يمكن وضعه في شكل فئات متساوية كما يتبين من الجدول التالي:

جدول (4ـ8) أعمار مجموعة من الأطفال

التكرار	الفئات
7	5 − 1
10	10 − 6
6	15 − 11
23	**المجموع**

ولحساب متوسط العمر لهؤلاء الأطفال ينبغي أولاً حساب مركز الفئة M وبعد ذلك تضرب التوزيعات بمركز الفئة:

$$\overline{X} = \frac{\sum FX}{N}$$

إن الإجراء الذي يمكن اتباعه لتطبيق هذه المعادلة:

1- حساب مركز الفئة لكل فئة من الفئات التي يحتويها الجدول.

2- ضرب مركز كل فئة في عدد تكراراتها.

3- جمع هذه النتائج.

4- تقسيم المجموع الكلي على عدد الحالات (N).

وتشير البيانات في الجدول التالي إلى مركز الفئة مضروباً في تكرار كل فئة من هذه الفئات.

جدول (4ـ9) حساب المتوسط لتكرارات مبوبة

الفئات (X)	مركز الفئة (M)	التكرار (F)	FM
5 – 1	3	7	21 = 7 × 3
10 – 6	8	10	80 = 10 × 8
15 – 11	13	6	78 = 6 × 13
المجموع		**N = 23**	$\sum FM = 179$

وبتعويض هذه البيانات في المعادلة، نتحصل على المتوسط العمري لهؤلاء الأطفال.

$$\overline{X} = \frac{\sum FX}{N} = \frac{179}{23}$$

$$= 7.8$$

الخصائص العامة للمتوسط والوسيط والمنوال:

يعد فهم الخصائص العامة للمنوال والوسيط والمتوسط أمراً مهماً وذلك لسببين أساسيين:

أولهما: يساعد فهم هذه الخصائص العامة لتفسير مقاييس النزعة المركزية. وثانيهما: يساعد على الإحساس بأنه باستطاعتنا الوصول إلى الاستدلال أو التعميم من العينات على المجتمع الذي سحبت منه هذه العينة.

الخصائص المؤثرة في التفسير:

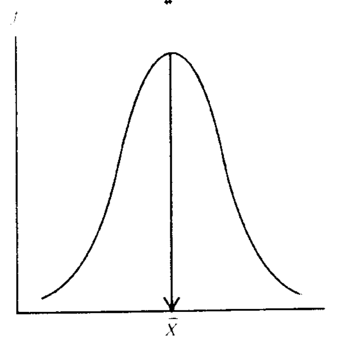
- **تأثير القيم المتطرفة:**

إن مقاييس النزعة المركزية المتمثلة في المنوال والوسيط أقل تأثيراً بالدرجات المتطرفة سواء أكانت هذه الدرجات عالية أم منخفضة، في حين على الجانب الآخر، أن المتوسط الحسابي يتأثر بشكل كبير بالقيم المتطرفة سواء كانت هذه القيم عالية أم منخفضة، وبناءً على ذلك يمكننا القول بأن المتوسط الحسابي يمكن النظر إليه بأنه مقياس غير ملائم للنزعة المركزية عندما يكون التوزيع ملتوياً التواءً شديداً كما يتضح ذلك من الشكل التالي [3]:

شكل (1-4) موقع المتوسط في توزيعات ملتوية وغير ملتوية

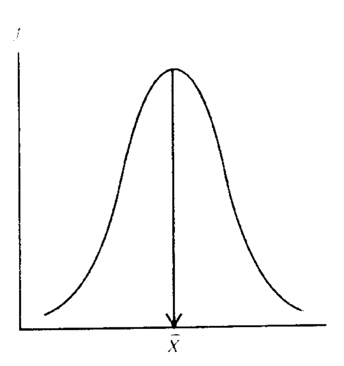
توزيع طبيعي غير ملتوٍ (أحادي)

(يكون المتوسط مؤشراً دقيقاً إلى حد نموذجي)



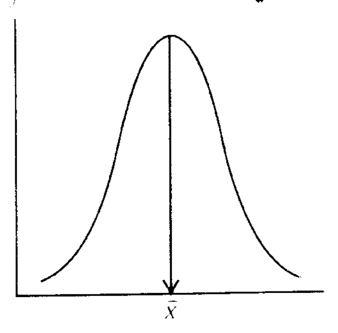
توزيع ملتوٍ (موجب)

(يكون المتوسط عالياً جداً ليمثل دقة النـزعة المركزية))

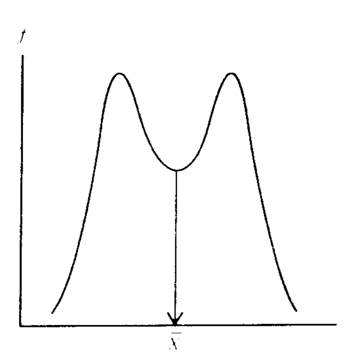


توزيع ملتوٍ (سالب)

(يكون المتوسط منخفضاً جداً ليعطي مؤشراً جيداً للنـزعة المركزية))



وفي أوقات أخرى عندما يعطي المتوسط إشارة مضللة للنزعة المركزية تشكل الدرجات نموذج توزيع مزدوج يقع حول النقطة حيث تقع قليل من الحالات. وأن مقداراً كبيراً من هذه الدرجات إما أن تكون عالية أو منخفضة إذا ما قورنت بالمتوسط الحسابي. ولذا كلما انحرفت الدرجات من النموذج الأحادي Unimodel يكون المتوسط عرضة للشك كمؤشر لمعدل الدرجة [4].



شكل (4-2): مركز المتوسط في توزيع ثنائي

وللتدليل على تأثر المتوسط بالقيم المتطرفة سواء أكانت هذه القيم عالية أو منخفضة نسوق المثال التالي:

إذا تعاملنا مع بيانات متعلقة بالعمر لعدد تسعة من الطلاب لبيان التغيرات التي تحدث:

الطلاب	الأعمار	الأعمار
1	19	19
2	20	20
3	20	20
4	20	20
5	21	21
6	22	22
7	22	22
8	23	23
9	24	42

فإذا غيرّنا عمر الطالب رقم 9 من 24 سنة إلى 42 سنة يمكننا أن نلاحظ أن المنوال والوسيط لم يتغيرا، في حين نلاحظ تغيرٍّ ملحوظ في المتوسط الحسابي لأعمار هؤلاء الطلاب. فقد بقي المنوال لهذه المجموعة 20 والوسيط 21 في حين وصل المتوسط الحسابي لهؤلاء الطلاب إلى 23.22 متأثراً بعمر الطالب الذي يبلغ من العمر 42 سنة.

تجدر الإشارة إلى أنه عندما نبين المتوسط للتوزيع ينبغي علينا أن ننظر إلى مدى القيم لنرى ما إذا قد تأثر المتوسط الحسابي بالدرجات المتطرفة إما إلى الطرف الأدنى أو الأعلى للتوزيع.

- **الخصائص المؤثرة في الاستدلالات:**

تعد الخصائص المتعلقة بالمتوسط الحسابي والوسيط والمنوال مهمة من خلال تشكيلها الأساس للإحصاء الذي ستتعلمه فيما بعد من خلال هذا الكتاب. فعند الحديث عن الإحصاءات الاستنتاجية التي من خلالها نسعى إلى عملية التعميم حول المجتمع من خلال العينات فإن المفاهيم التي سنتناولها في هذا الجزء تساعدنا في فهم المفاهيم الأساسية للإحصاءات الاستدلالية.

- **الثبات في سحب العينات العشوائية:**

يعتبر المتوسط الحسابي أكثر المقاييس ثباتاً في مقاييس النزعة المركزية للعينات المسحوبة بشكل عشوائي. فإذا أراد الباحث سحب مجموعة من العينات من نفس المجتمع فإنه سيواجه أقل تباينات في المتوسط إذا قورنَ ذلك بالوسيطات أو المنوال. دعنا نورد المثال التالي للتدليل على ذلك:

العينات	المتوسط	الوسيط	المنوال
العينة 1	46.20	40	35
العينة 2	50.62	44	40
العينة 3	46.68	46	28
العينة 4	46.30	47	51
العينة 5	47.72	45	45

المصدر J. Richard Kendrick Social Statistics: AN Introduction Using SPSS, USA, 2005, P. 158

ما نلاحظه من خلال هذه البيانات أن مدى القيم التي نتحصل عليها من كل واحد من مقاييس النزعة المركزية فإن متوسطات المدى من الأدنى 20. 46 سنة إلى الأعلى 62. 50 سنة بفارق يصل إلى 4.42 ومدى الوسيطات من الأدنى 40 سنة إلى الأعلى 47 سنة بفارق يصل إلى 7 سنوات. في حين أن مدى المنوال يبدأ من 28 سنة إلى 21 سنة بفارق يصل إلى 23 سنة. من هذا العرض يتبين لنا أن متوسطات العينات تبرز أقل التباينات عندما تقاس بمدى القيم.

من خلال هذه العمليات الحسابية البسيطة، يمكننا ملاحظة أن المتوسطات التي تم الحصول عليها من العينة قريبة من المتوسط الفعلي للمجتمع.

في حقيقة الأمر أن متوسطات العينات تكون قريبة من متوسط المجتمع الذي سحبت منه هذه العينات أكثر منه في حالة الوسيط والمنوال عندما يتعلق الأمر بالوسيط والمنوال والمجتمع فإذا كان متوسط الأعمار لدى الأفراد في المجتمع يصل إلى 28. 46 فبالنظر إلى هذا المتوسط في إطار شكل واحد من متوسطات العينة نجد أنَّ:

العينات	متوسط العينة	متوسط المجتمع	الفرق
العينة 1	46.20	46.28	−0.08
العينة 2	50.62	46.28	4.34
العينة 3	46.98	46.28	0.70
العينة 4	46.30	46.28	0.20
العينة 5	47.72	46.28	1.44

المصدر : J. Richard Kendrick, Social Statistics: Ibid, P. 158

إذا أضفنا الفروق بين كل متوسطات العينة ومتوسط المجتمع متجاهلين الإشارات (أي التعامل مع الأرقام السلبية كما لو كانت موجبة)، فإن مجموع الفروق يصل إلى 6.58 وإذا قسمنا مجموع الفروق على 5 (عدد متوسطات العينات) فإننا سنتحصل على متوسط الفروق يصل إلى 1.32، متوسط الفروق بين متوسطات العينات ومتوسط المجتمع يزيد قليلاً عن سنة.

دعنا هذه المرة نقوم بعملية حسابية تتعلق بوسيطات العينات ومنوال العينات والسؤال الذي يمكن طرحه هو، هل بإمكاننا أن نتحصل على نتائج متشابهة؟ إذا أضفنا الفروق بين كل واحد من وسيطات العينات ووسيط المجتمع يمكننا الحصول على الآتي:

العينات	وسيط العينة	وسيط المجتمع	الفرق
عينة 1	40.00	44	−4
عينة 2	44.00	44	0
عينة 3	46.00	44	2
عينة 4	47.00	44	3
عينة 5	45.00	44	1

المصدر: Ibid, P. 158.

مجموع الفروق الكلية (متجاهلين الإشارات) مجموع الفروق بين وسيط العينات ووسيطات العينات يساوي 10 وبتقسيم هذا المجموع على 5 (عدد العينات) يتضح لنا أن متوسط الفروق بين وسيطات العينات ووسيط المجتمع يصل إلى سنتين.

تجدر الإشارة إلى أنه يمكننا عمل الشيء نفسه فيما يتعلق بالمنوال. فإذا كان منوال المجتمع يساوي 33 سنة فإذا أضفنا الفروق بين كل منوال متعلق بالعينات ومنوال المجتمع متجاهلين الإشارات فإننا سنتحصل على فرق يصل إلى 44 سنة وإذا قسمنا هذا العدد على 5 (عدد العينات) فإن متوسط الفروق بين منوال العينات ومنوال المجتمع يصل إلى 8.8 سنة.

العينات	منوال العينة	منوال المجتمع	الفرق
عينة 1	33	35	2
عينة 2	33	40	7
عينة 3	33	28	5
عينة 4	33	51	18
عينة 5	33	45	12

المصدر: Ibid, P. 158.

من خلال هذه الأمثلة تم سحب متوسطات عينات من مجتمع معروف المعالم مثل المتوسط الحسابي، والوسيط والمنوال. ومن خلال هذا الكتاب في فصوله اللاحقة فإن هذه الخاصية المتعلقة بالمتوسط الحسابي - الثبات عبر سحب العينات عشوائياً من المجتمع - يستخدم هذا المتوسط للتقديرات حول المجتمعات حتى وإنْ لم يكن شيئاً معروفاً حول المجتمع نفسه.

- **مجموع الانحرافات عن المتوسط الحسابي تساوي صفرا (0):**

من الخصائص الأخرى للمتوسط الحسابي أن مجموع الانحرافات عن المتوسط الحسابي دائماً تساوي صفراً وهذا يعني أنه إذا طرحنا كل درجة من مجموع بيانات المتوسط الحسابي لكل الدرجات فإننا سنتحصل دائماً على صفر والمثال التالي يوضح ذلك:

الدرجة	19	20	20	21	21	21	22	22	23
المتوسط	21	21	21	21	21	21	21	21	21
الانحرافات عن المتوسط	−2	−1	−1	0	0	0	1	1	2
مجموع الانحرافات عن المتوسط تساوي: صفراً									

المصدر: Ibid, P. 159.

اختيار مقاييس النزعة المركزية:

لاختيار مقاييس النزعة المركزية، ينبغي على الباحث أن يكون على دراية كاملة بمستوى القياس للمتغير تحت الدراسة. وهذه المعرفة بمستوى القياس بشكل عام، تساعد الباحث فيما إذا كان ينبغي عليه إعداد تقريرٍ متضمنٍ للمنوال، أو الوسيط أو المتوسط الحسابي. والدليل التالي يوضح عملية اختيار مقاييس النزعة المركزية:

1- المنوال: يستخدم المنوال في الأحوال التالية:

- عندما يقاس المتغير على المستوى الاسمي.
- عندما يريد الباحث قياساً سريعاً وسهلاً للمتغيرات الترتيبية، والمتغيرات ذات المسافات المتساوية والنسبية.
- عندما يريد أن يقدم تقريراً للدرجات الأكثر شيوعاً.

2- الوسيط: يستخدم الوسيط:

- عندما يقاس المتغير على المستوى الترتيبي.
- وعندما يقاس المتغير على المستوى ذي المسافات والنسبي ويمتلك توزيعاً عالياً من الالتواء.

3- المتوسط الحسابي: يستخدم المتوسط الحسابي:

- عندما يكون المتغير قد تم قياسه على مستوى المقياس ذي المسافات والنسبي (ما عدا أنه عندما يكون المتغير على درجة عالية من الالتواء).
- عندما يريد الباحث أن يقدم درجة نموذجية. فالمتوسط الحسابي يعتبر نقطة ارتكاز يعادل بشكل دقيق كل الدرجات.
- إذا كان الباحث يتوقع تحليلات إحصائية إضافية.

المصدر: Joseph F. Healey, The Essentials of Statistics: A TOOL for Social Research, .Wadsworth Cengage Learning, USA, 2010, P. 95

استخدام مقاييس النزعة المركزية:

الشكل التالي يوضح لنا التطبيقات الصحيحة لكل من المنوال، والوسيط، والمتوسط، وأن استخدامات هذه التقنيات الإحصائية يعتمد بالدرجة الأولى على مستوى قياس المتغير الذي نسعى إلى تطبيق هذه الإحصاءات عليه. ويمكن للباحث أن يستعين بهذا الشكل كمرشد يساعده في كيفية اختيار كل واحد من تقنيات النزعة المركزية:

المتغيرات العددية:

- متغيرات ذات مسافات

أو

نسبية ↓

(يستخدم المنوال، والوسيط والمتوسط)

↓

ولإيجاد المتوسط يدوياً: استخدم المعادلة المتعلقة بالبيانات الخام، أو إذا كانت البيانات في شكل توزيعات تكرارية: استخدم المعادلة المتعلقة بذلك.

↓

أما إذا أردت إيجاد المتوسط باستخدام البرنامج SPSS عليك اتباع الأمر التالي:

⇨ Analyze
⇨ Descriptive Statistics
⇨ Frequencies
Click on the Statistics button

انقر على زر الإحصاء

Check Mean

وضع، إشارة أمام المتوسط

المتغيرات المقولية:

أ- متغيرات اسمية: يتم استخدام المنوال ولإيجاد المنوال يدوياً، انظر إلى القيمة الأعلى في التوزيع التكراري.

أما إذا أردت إيجاد المنوال باستخدام برنامج Spss، عليك اتباع الأمر التالي:

⇨ Analyze
⇨ Descriptive Statistics
⇨ Frequencies
Click on the Statistics button
Check Mode :

(انقر على زر الإحصاء وضع إشارة أمام المنوال).

ب- بيانات ترتيبية يتم استخدام المنوال والوسيط ولإيجاد الوسيط يدوياً، أوجد الحالة الوسيطية باستخدام المعادلة المتعلقة بذلك، وبعد ذلك استخدم التوزيع التكراري لمعرفة قيمة الوسيط.

أما إذا أردت إيجاد الوسيط باستخدام برنامج Spss عليك اتباع الأمر التالي.

⇨ Analyze
⇨ Descriptive Statistics
⇨ Frequencies
Click on the Statistics button
Check Mediam :

(انقر على زر الإحصاء، وضع إشارة أمام الوسيط)

ثانياً: مقاييس التشتت:

في هذا الجزء سنتناول بالتفصيل استخدامات الانحرافات عن المتوسط الحسابي لحساب المقاييس التي من خلالها يمكننا تقييم توزيعات التشتت. إذا كانت مقاييس النزعة المركزية تكشف لنا مجموعة واحدة من قيم البيانات حول توزيع الإجابات واتجاه تجمعها فإن مقاييس التشتت تكشف لنا أمراً آخر وهو، إلى أي مدى قد انتشرت أو تشتت، أو تباينت تلك التوزيعات؟. وتجدر ملاحظة أنّ مقاييس النزعة المركزية ومقاييس التشتت يعملان معاً. فمقاييس النزعة المركزية هي إحصاءات وصفية تشير إلى قيمة المعدل في التوزيع في حين تشير مقاييس التشتت إلى تباين الدرجات في التوزيع.

يرتبط تحليل التشتت بشكل أساسي بتحليل التوزيعات التكرارية تلك التوزيعات التي تساعد الباحث في تقييم مدى التجانس (التشابه) والتّغاير (الفرق) في توزيعات إجابات المبحوثين على متغيرات معينة، إذاً مقاييس التشتت تمدنا بالإحصاء لتقييم إلى أي مدى يكون التوزيع متجانساً أو متغايراً؟ والسؤال الذي يمكن طرحه هنا: لماذا يعتبر فهم تشتت إجابات المبحوثين أمر مهم لدى الباحث؟ للإجابة على هذا السؤال نورد جملة من العوامل:

- تساعد مقاييس التشتت على فهم مفهوم المتغيرّية Variability حيث تشير المتغيرّية إلى درجة التباين في الإجابات المتعلقة بالمتغير المدروس فعندما تكون توزيعات الإجابة أكثر تجانساً يصاحبها قليل من المتغيرّية (التباين) في إجابات المبحوثين حول المتغير المدروس والعكس بالعكس. كلما كانت التوزيعات أكثر تغايراً كانت نسبة التباين عالية في إجابات المبحوثين وكما نرى لاحقاً في مَتْنِ هذا الكتاب، فإن مفهوم المتغيرّية يلعب دوراً كبيراً في الإحصاءات الاستنتاجية، وفي قدرة الباحث على التعميم من العينات العشوائية التي يتم سحبها من المجتمع.

دعنا الآن نورد المثال التالي لتوزيعين من الحالات طبقاً للدخل السنوي كما هو مبين في الجدول التالي:

جدول (4ـ10) الدخول السنوية

مجموعة (1) $	مجموعة (2) $
5000	20,000
6500	28,500
8000	35,000
55,000	36,000
85,000	40,000

المصدر: George Argyrous, op. Cit., P.73.

إن المتوسط الحسابي لهاتين المجموعتين يكون متساوياً:

$$\overline{X_1} = \frac{5000 + 6500 + 8000 + 55{,}000 + 85{,}000}{5} = \$\ 31.900$$

$$\overline{X_2} = \frac{20{,}000 + 28{,}000 + 35{,}000 + 36{,}000 + 40{,}000}{5} = \$\ 31.900$$

إن هذه التوزيعات لديها نفس المتوسط الحسابي ومع ذلك فإنه من الواضح أن نجد أيضاً فرقاً جوهريًّاً بين التوزيعات. وبالرغم من أن المتوسط الحسابي لهاتين المجموعتين متساوٍ إلا أن هناك فرقاً كبيراً في تشتت الدرجات.

دعنا نبدأ بمقاييس التشتت للبيانات المقاسة على المستوى ذي المسافات والنسبي: المدى، المدى الربيعي، الانحراف ومُعَامِل التغاير النسبي، ثم بعد ذلك نقوم بدراسة مقاييس التشتت للبيانات المقولية: مؤشر التباين النوعي.

المدى:

إن أبسط مقاييس التشتت هو المدى وذلك لسهولة حسابه. ويُعرف المدى: بأنه الفرق بين أكبر قيمة وأصغر قيمة في التوزيع فإذا رجعنا إلى المثال السابق جدول (1) المتعلق بتوزيع الدخل لمجموعتين:

فالمدى للمجموعة (1): $R_1 = 85,000 - 5000 = 80,000$

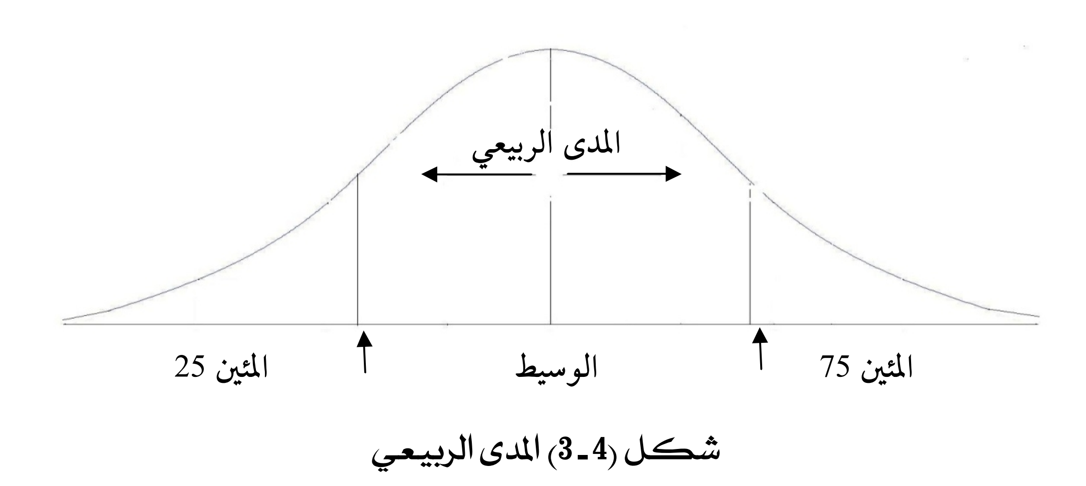
والمدى للمجموعة (2): $R_2 = 40,000 - 20,000 = 20,000$

والشيء الذي يمكن ملاحظته من خلال حساب المدى لهاتين المجموعتين هو أنه بالرغم من أن التوزيعين لديهما نفس المتوسط الحسابي، إلا أن الفرق واضح في توزيع الدرجات حول المتوسط. فالمجموعة (1) لديها تباين أكثر إذا ما قورنت بالمجموعة (2).

إن أهمية المدى كمقياس للتشتت كما ذكرنا تكمن في سهولة حسابه وذلك بطرح أصغر قيمة من أكبر قيمة ومن هنا نجد أنه بالرغم من هذه الأهمية فإنها تقلل من قيمته حيث إنه يعتمد على القيمتين المتطرفتين فإذا نظرنا إلى توزيع الدخل للمجموعة (2) نجد أن كل الحالات تقع في مدى 20,000 بين 20,000 و 40,000 فإذا أضفنا دخل شخص سادس يقدر بـِ 150,000 لهذه المجموعة ففي هذه الحالة يمتد المدى ليصبح 130,000 ولتعويض التأثير يستخدم المدى الربيعي Interquartile Range (IQR) وهو عبارة عن الفرق بين الربيع الثالث والربيع الأول أي: $Q_3 - Q_1$

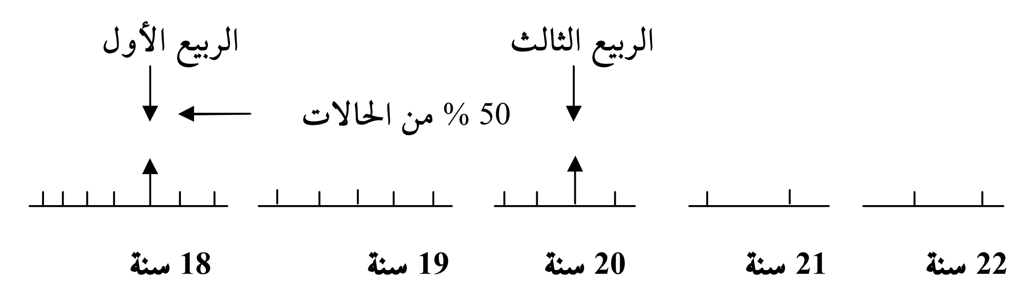
المدى الربيعي:

إنه من خلال المدى الربيعي يمكننا التغلب على المشكلات التي يمكن أن تظهر مع المدى البسيط وذلك من خلال تجاهل الدرجات المتطرفة لأي توزيع. ويُعرف المدى الربيعي: بأنه الفرق بين الحدود العليا للربيع الأول والربيع الثالث بمعنى آخر هو المدى لمنتصف 50 % من الحالات في سلسلة مرتبة انظر شكل (1).



شكل (4-3) المدى الربيعي

ولحساب المدى الربيعي يمكننا استخدام المثال التالي: إذا كان لدينا 20 مبحوثاً أي 20 حالة ولذلك فإن كل مئين يحتوي على 20 ÷ 4 = 5 حالات. إن نهاية المئين الأول تنتهي بشخص عمره 18 سنة ونهاية المئين الثالث تنتهي بشخص 15 يبلغ من العمر 20 سنة. انظر الشكل التالي:



المصدر: George Argyrous, Statistics for Social & Health Research, with a Guide to SPSS، SAGE Publications, London ,2001, p.75.

وبالتالي يكون المدى الربيعي 20 – 18 = 2 وإذا قارنا المدى الربيعي بالمدى البسيط فإن المدى الربيعي لن يتغير بشكل جذري إذا ما أضفنا شخصاً أو شخصين سواء أكانا كبيرين في السن أم صغيرين لأي جانب من التوزيع.

الانحراف المعياري:

يعتبر الانحراف المعياري أحد مقاييس التشتت ويحدد الانتشار من خلال بيان الفرق بين كل درجة من الدرجات في التوزيع ومتوسطها؛ أي انحراف القيمة عن متوسطهما الحسابي $\sum(X-\overline{X})$

وعند حساب الانحراف المعياري تستخدم المعادلات وفقاً للبيانات المعطاة سواء أكانت بيانات خام أم مبوبة. وفي كل حالة من هاتين الحالتين يستخدم الرمز الجبري S للإشارة للانحراف المعياري للعينة و σ للانحراف المعياري للمجتمع وعليه فإن المعادلة المستخدمة لإيجاد الانحراف المعياري للبيانات الخام في حالة العينة:

$$S=\sqrt{\frac{\sum(X-\overline{X})^2}{N}}$$

في حين أن المعادلة المستخدمة لإيجاد الانحراف المعياري للبيانات الخام في المجتمع:

$$\sigma=\sqrt{\frac{\sum(X-u)^2}{N}}$$

وبنظرة فاحصة لهاتين المعادلتين نجدهما يشيران إلى فكرة أن الانحراف المعياري هو مسافة معدل Average Distance لكل درجة من المتوسط الحسابي. فالبسط في المعادلة هو ببساطة الفرق في أعداد المشاهدات.

وللتركيز على فكرة الانحراف المعياري بشكل دقيق، يمكننا الرجوع إلى التوزيع العمري لمجموعة العشرين فرداً والتي تم حساب متوسطها الحسابي المقدر بـِ 19.3 سنة، حيث وجدنا أن كل الدرجات تنحرف عن متوسطها الحسابي إما سلباً (درجات تحت المتوسط) أو إيجاباً (درجات فوق المتوسط).

تجدر الإشارة هنا إلى أنه ليس بإمكاننا إضافة كل الانحرافات الموجبة (درجات فوق المتوسط) مع كل الانحرافات السالبة (درجات تحت المتوسط) حيث إنه بالتعريف أن المجموع يساوي صفراً (0) وبهذا فإن معادلة الانحراف المعياري تربع الفروق، وعليه

تتحول كل الانحرافات إلى أرقام موجبة. وبالتالي كلما كانت الفروق كبيرة كانت قيم الانحراف المعياري أكبر.

مثال:

دعنا الآن عملياً نقوم بحساب الانحراف المعياري لهذا التوزيع حيث أنه بإمكاننا استخدام المعادلة أعلاه حيث أوردنا هذه المعادلة لكي نبين الفكرة بأن الانحراف المعياري هو متوسط المسافة من المتوسط الحسابي. ومن الناحية الفعلية إذا أردنا حساب الانحراف المعياري لبيانات خام فإننا سوف نتبنى معادلة تختلف قليلاً ولكنها سهلة الاستخدام وتعطي نفس النتيجة التي تعطيها المعادلة الأولى:

$$S=\sqrt{\frac{\Sigma X^2-\frac{(\Sigma X)^2}{N}}{N-1}}$$

حيث إن: ΣX^2 مجموع كل مربع الدرجات

$(\Sigma X)^2$ مجموع كل الدرجات المربعة

إن الخطوة الأولى هي تربيع كل درجة من الدرجات، وبعد ذلك تضاف إلى بعضها البعض. أما الخطوة الثانية فهي تعكس الإجراء أي تضاف الدرجات لبعضها البعض ثم يربع المجموع.

والجدول (4 – 11) يوضح هذا الإجراء:

جدول (4ـ11) إجراء حساب الانحراف المعياري لبيانات خام (العمر)

الحالة	العمر بالسنوات X^2	X^2
1	18	324
2	21	441
3	20	400
4	18	324
5	19	361
6	18	324
7	22	484
8	19	361
9	18	324
10	20	400
11	18	324
12	19	361
13	22	484
14	19	361
15	20	400
16	18	324
17	21	441
18	19	361
19	18	324
20	20	400
	$\Sigma X = 387$	$\Sigma X^2 = 7523$

وبتعويض هذه الأرقام بالمعادلة نتحصل على:

$$S=\sqrt{\frac{\Sigma X^2-\frac{(\Sigma X)^2}{N}}{N-1}}=\sqrt{\frac{7523-\frac{(387)^2}{20}}{20-1}}=$$

$$= 1.35 \text{ سنة}$$

معامل التباين النسبي (CRV):

تجدر الإشارة إلى أن هناك بعض القصور الذي يمكن التغلب عليه عند استخدام الانحراف المعياري وذلك من خلال استخدام معامل التباين النسبي CRV ويستخدم معامل التباين النسبي في:

1- مقارنة التوزيعات التي تم قياسها على نفس الوحدات بمتوسطات مختلفة.

2- مقارنة التوزيعات التي تم قياسها على وحدات مختلفة.

إنه لا يمكننا القول فيما إذا كان المتوسط العمري للمثال السابق 1,35 سنة كمية صغيرة أم كبيرة للتشتت حول المتوسط. علاوة على ذلك فإن الانحراف المعياري لمجموعة واحدة من المشاهدات لا يمكن مقارنتها بمجموعة أخرى من الدرجات لنقرر أيًا من التوزيعين أكثر تشتتاً. فعلى سبيل المثال، قد يكون لدى التوزيعين نفس الانحراف المعياري 1,35 إلا أن المتوسط لكل منهما 5 سنوات و 50 سنة على التوالي، فإنه من الوضوح أن الانحراف المعياري 1.35 يمثل نسبياً أكثر تبايناً للتوزيع الذي يعكس متوسطاً أصغر [5].

في موقع آخر، قد يرغب الباحث بمقارنة التباين في متغيرين منفصلين ثم قياسهما بوحدات مختلفة، فعلى سبيل المثال، قد يرغب الباحث في معرفة ما إذا كان عمر مجموعة من المبحوثين بانحراف معياري 1,35 سنة تصور لنا أكثر تبايناً من دخلهما الأسبوعي بانحراف معياري 65 دولار. في هذه الحالة لا نستطيع مقارنة هذين الانحرافين المعياريين ونقول إن أحد المتغيرات أكثر تشتتاً من الآخر لأن كل واحد من هذين المتغيرين تم قياسه بوحدات مختلفة فكأننا في هذه الحالة، وبشكل فعال نقارن التفاح بالبرتقال.

ولكي نعطي مقياساً معيارياً للتشتت ينبغي علينا أن نقوم بعملية حساب معامل التباين النسبي CRV التي تعكس الانحراف المعياري كنسبة مئوية للمتوسط [(6)].

$$CRV = \frac{S}{\overline{\overline{X}}} \times 100$$

وباستخدامنا لهذه المعادلة لتوزيع الأعمار لعدد 20 من المبحوثين نتحصل على:

$$CRV = \frac{S}{\overline{\overline{X}}} \times 100 = \frac{1.35}{19.25} \times 100$$

$$= 7\ \%$$

إذا كان لدينا مجموعة أخرى من الناس بمعلومات كافية عن أعمارهم عندئذ يمكننا حساب CRV لهذه المجموعة ومقارنتها بالمجموعة الأولى لنرى أيًّا من هاتين المجموعتين لديها كمية كبيرة من التشتت.

وإذا تكشف لنا أن المجموعة الثانية من المبحوثين لديهم انحراف معياري لأعمارهم يصل إلى خمس سنوات ومتوسط عمري يصل إلى 21 سنة فإن قيمة CRV ستكون:

$$CRV = \frac{S}{\overline{\overline{X}}} \times 100 = \frac{5}{21} \times 100$$

$$= 24\ \%$$

إن المجموعة الثانية من الناس تعطي أكثر تبايناً في أعمارها إذا ما قورنت بالمجموعة الأولى. ومن هنا يمكننا القول إن المجموعة الثانية تظهر أكثر تبايناً من المجموعة الأولى حيث إن هذا التباين يصل إلى فرق يقدر بـِ 17 %.

مؤشر التباين النوعي (IQV):

لقد بيّنا أن مقاييس التشتت يمكن تطبيقها على تلك البيانات المقاسة على المستوى ذي المسافات والنسبي لأن ذلك يتطلب قياس المسافة بين الدرجات.

إن السؤال الذي يمكن طرحه هنا: كيف يمكننا أن نبين كمية التباين في توزيع تكون

البيانات فيه بيانات مقولية فقط؟ كثيراً ما يهمل مقياس التباين في مثل هذه الحالات الملائمة ويعوض عنه بمؤشر التباين النوعي IQV. ويعرف مؤشر التباين النوعي IQV بأنه عدد الفروق بين الدرجات في التوزيع تعرض كنسبة لعدد الفروق الكلية الممكنة وتسمح لنا IQV باستخدام كمية التباين التي يتضمنها التوزيع حتى عندما تكون لدينا بيانات اسمية. فعلى سبيل المثال، لماذا كانت لدينا بيانات حول متغير النوع (اسمي) والتي يصعب حصولنا على التباين بأي مقياس من مقاييس التشتت. انظر البيانات التالية.

جدول (4 ـ 12) نوع المبحوثين

النوع	التكرار
ذكر	12
أنثى	8
المجموع	20

فإذا كانت كل الحالات ذكوراً كانت أو إناثاً لديها نفس الدرجة على المتغير المدروس، ففي هذه الحالة لا يوجد تباين. وبالتالي تشكل هذه البيانات كمية أدنى من التباين الممكن مشاهدته.

إن أقصى تباين محتمل الذي ربما نستطيع مشاهدته في التوزيع هو: إذا كانت الحالات قد تم توزيعها بالتساوي عبر فئات المتغير كما هو الحال في المثال التالي جدول (4 – 13).

جدول (4 ـ 13) نوع المبحوثين: أقصى تباين محتمل

النوع	التكرار
ذكور	10
إناث	10
المجموع	**20**

ففي هذا التوزيع لدينا مئة من الفروق كل 10 إناث تختلف عن كل واحد من 10 من الذكور طبقاً للنوع.

ثمة طريقة بسيطة لحساب أقصى تباين محتمل من الفروق التي يكون باستطاعتنا مشاهدتها لأي مجموعة من البيانات المقولية وذلك باستخدام المعادلة التالية:

$$\text{أقصى فروق محتملة} = \frac{N^2(K-1)}{2K}$$

حيث إن: K عدد الفئات.

ويمكننا استخدام هذه المعادلة للوصول إلى أقصى عدد من الفروق لعدد الحالات والفئات للمثال الذي بين أيدينا.

$$\text{أقصى فروق محتملة} = \frac{N^2(K-1)}{2K} = \frac{20^2(2-1)}{2(2)}$$

$$= 100$$

إذا أمعنا النظر في البيانات الفعلية الواردة في جدول (3) فإنه من الواضح أن هذه البيانات تشبه بشكل قريب أقصى تباين كما ورد في بيانات جدول (4) أكثر من الموقف الذي لا يوجد فيه تباين. إن IQV يسمح لنا بأن نعبر عن ذلك كمياً ولفعل ذلك فإن الأمر يتطلب أن نحدد عدد الفروق المشاهدة في توزيع الدرجات التي نقوم بتحليلها، فإذا أخذنا 1 من 8 إناث، فالسؤال الذي يمكن طرحه هو كم من الأفراد الآخرين في التوزيع يكونون مختلفين وفقاً للنوع؟ من الواضح أن لدينا 12 حالة من الذكور في التوزيع ولكل واحدة من 8 من الإناث سوف يكون هناك 12 من حالة الذكور الآخرين في التوزيع الذين هم مختلفون عن النساء مولّدين عدداً كلياً يصل إلى 96 من الفروق المشاهدة. وعليه فإن IQV لنوع المبحوثين سيكون:

$$IQV = \frac{\text{الفروق المشاهدة}}{\text{أقصى فروق محتملة}} = \frac{96}{100}$$

$$= 0.96$$

ويشير IQV إلى 0.96 وإلى أن هناك كميةً كبيرةً من التباين في البيانات المتعلقة بهذا المتغير ولكن على الجانب الآخر، فإنه إذا كان لدينا كل الإناث أو كل الذكور وبالتالي لا يوجد لدينا فروق مشاهدة في البيانات من هنا، ومن السهولة بمكان، أن نرى وبشكل نسبي أن قيمة IQV سوف تكون مساوية لصفر (0) مشيرة إلى أنه لا يوجد تباين [7]. ولزيادة التوضيح دعنا نسوق مثالاً آخر لحساب كمية التباين مستخدمين نفس المقياس IQV ويدور هذا المثال حول توزيع المبحوثين طبقاً للرضا الصحي.

جدول (4-14) درجة الرضا الصحي للمبحوثين

الرضا الصحي	التكرار
غير راضٍ	7
راضٍ	5
راضٍ جداً	8
المجموع	**20**

إن السؤال الذي يمكن طرحه هنا هو كم عدد المرات التي تختلف فيها هذه الحالات - في هذا التوزيع - عن الحالات الأخرى، دعنا نبدأ بالحالات التي أجابت بأنها راضية جداً (8) نجد أن كل واحد من هؤلاء يختلف عن أولئك الذين أجابوا غير راض (7) وراض (5) مولّدة (96) من الفروق وبهذا يمكننا إضافة 35 من الفروق (الذين أجابوا أنهم غير راضين 7) (والذين أجابوا أنهم راضون 5).

ومن هنا نجد أن العدد الكلي للفروق المشاهدة تكون كالتالي:

الفروق المشاهدة = 8 (5) + 8 (7) + 5 (7)

= 131

إن الحد الأقصى لعدد الفروق التي باستطاعتنا مشاهدتها (إذا ما توزعت الحالات بشكل متساوٍ عبر الفئات الثلاثة) هو:

$$\text{أقصى فروق محتملة} = \frac{N^2(K-1)}{2K} = \frac{20^2(3-1)}{2(3)}$$

$$= 133.3$$

وعليه فإن قيمة IQV تساوي:

$$IQV = \frac{\text{الفروق المشاهدة}}{\text{أقصى فروق محتملة}} = \frac{131}{133.3}$$

$$= 0.98$$

وتشير هذه الدرجة إلى أنَّ هناك تقريباً أقصى تباين محتمل بين هذه الحالات فيما يتعلق بدرجة الرضا الصحي. كما يمكننا أيضاً القول في هذا السياق بأنّ هناك تقريباً نفس كمية التباين لهذه الحالات فيما يتعلق بدرجة الرضا مثل كمية التباين فيما يتعلق بالنوع(8).

ولزيادة التوضيح نورد المثال التالي:

مثال: زيارة الأقارب:

عدد مرات الزيارة	التكرار	%
مرة في الأسبوع	39	52.0
مرة في الشهر	11	14.7
مرة في العام	4	5.3
لا زيارة على الإطلاق	21	28.0
المجموع	**75**	**% 100.0**

الخطوة الأولى: ضرب التكرار الأول (في الأسبوع) 39 في مجموع التكرارات الموالية:

$$FF = (39)(11) + (4) + 21$$
$$= (39)(36)$$
$$= 1,404$$

الخطوة الثانية: ضرب التكرار الثاني (مرة في الشهر) 11 في التكرار بين المواليين:

$$FF = (11)(4) + 21$$
$$= (11)\ (25)$$
$$= 275$$

الخطوة الثالثة: ضرب التكرار الثالث في التكرار الذي يليه مباشرة:

$$FF = (4)\ (21)$$
$$= 84$$

الخطوة الرابعة: إضافة كل FF_s للحصول على الإجمالي ($\sum$):

$$FF = 1{,}404 + 275 + 84$$
$$= 1{,}763$$

والآن يمكننا بيان الخطوات التي نتحصل فيها على: أقصى فروق محتملة:

$$\text{أقصى فرق محتمل} = \frac{N^2(K-1)}{2K}$$

$$= \frac{75^2(4-1)}{2(4)}$$

$$= 2{,}109375$$

$$IQV = \frac{\text{الفروق المشاهدة}}{\text{أقصى فروق محتملة}}$$

$$= \frac{1.763}{2{,}109{,}375}$$

$$IQV = .84$$

ولتفسير IQV، يمكننا القول بأنّ قيمتّها دائماً قريبة من 1.0 أكثر من (0). وتشير IQV إلى أن التوزيع تماماً متغاير الخواص، وعند التعامل مع الإحصاء، فإننا دائماً نتساءل ما إذا كانت الإجابات التي تقدمها إجابات منطقية. وهل هذه الإجابات تؤيد من خلال طرق أخرى ننظر بها للمتغير؟ هل التوزيع التكراري في هذه الحالة، يدعم هذا المطلب؟

إن الإجابة على ذلك يمكن أن تكون نعم ولا. وبالرغم من وجود انتشار عبر الفئات هذا المتغير = 52 % من المبحوثين يتركزون في فئة الذين أجابوا (مرة في الأسبوع). وقرابة 15 % من المبحوثين يقومون بزيارة الأقارب مرة في الشهر. وحوالي أكثر من ثلث المبحوثين لا يزورون أقاربهم أو يقومون بالزيارة مرة في العام. إن الإجابات المتعلقة بزيارة الأقارب أكثر تجانساً إذا ما قورنت بمؤشر IQV.

تجدر الإشارة إلى أن برنامج SPSS لا يحتوي على هذه الإحصاءات مثل IQV لتحليل المتغيّرية للمتغيرات المقاسة على المستوى الاسمي (من المحتمل كتابة معادلة للبرنامج لكي يقوم بحساب IQV). ومن هنا نجد أن هناك عدة إحصاءات لتحليل المتغيّرية بين المتغيرات العددية والترتيبية مثل: المدى، والمدى الربيعي يمكن استخدامها [9].

استخدام مقاييس التشتت:

الإطار التالي يوضح لنا التطبيقات الصحيحة لمقاييس التشتت. وهناك مجموعة من الإحصاءات الملائمة المعتمدة على مستوى قياس المتغير الخاضع للتقييم. ويمكن للباحث أن يستعين بهذا الإطار لمساعدته متى وكيف يستخدم هذه الإحصاءات بشكل صحيح؟

إطار:

المتغيرات العددية:

- متغيرات ذات مسافات متساوية ونسبية (يتم استخدام المدى، والمدى الربيعي، والانحراف المعياري)
- من خلال العملية اليدوية يمكن الحصول على الانحراف المعياري للبيانات الخام باستخدام المعادلة:

$$S = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N-1}}$$

- أما إذا كانت البيانات منظمة في توزيع تكراري. استخدم المعادلة للتوزيع التكراري غير المبوب:

$$S = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N-1}}$$

- أما إذا أراد الباحث إيجاد الانحراف المعياري باستخدام برنامج SPSS إتباع الأمر التالي:

⇨ Analyze
⇨ Descriptive Statistics
⇨ Frequencies
Click on the Statistics button

أنقر على زر الإحصاء

Check Range, quartiles, and St. Deviation.

(ضع إشارة أمام المدى، الربيعات والانحراف المعياري)

المتغيرات المقولية:

أ- متغيرات اسمية: (يتم استخدام مؤشر التباين النوعي IQV) وبالتالي يمكن حسابه يدوياً باستخدام المعادلة التالية:

$$IQV = \frac{\text{مجموع الفروق المشاهدة}}{\text{أقصى فروق محتملة}}$$

مجموع الفروق المشاهدة = $\Sigma(F)(F)$

ونعني بـ Σ الجمع التكرارات التالية وأن FF تعني ضرب التكرار في الفئة المعطاة بالتكرارات الموالية للفئات.

- إضافة كل ff_s للحصول على الإجمالي Σ
- أقصى فروق محتملة: $\frac{N^2(K-1)}{2K}$
- أما إذا أراد الباحث إيجاد IQV باستخدام برنامج Spss فإن هذا البرنامج لا يتضمن إحصاءات متعلقة بالمتغيرات الاسمية ومع ذلك فإن تحليل التوزيع التكراري للمتغيرات الاسمية يمكن أن تساعد في تقييم المتغيرّية.

ب- متغيرات ترتيبية:

- يتم استخدام المدى، والمدى الربيعي.
- من خلال العملية اليدوية يمكن إيجاد، المدى الربيعي (IQR) باستخدام المعادلات:
- إيجاد المدى الربيعي $IQR = Q_1 - Q_3$
- إيجاد الحالة في الربيع الأول = .25 (N+1) أو تقسيمها (رياضياً النتيجة واحدة).
- ولإيجاد الربيع الثالث: نستخدم المعادلات التالية: (N+1) X.75
- أما إذا أراد الباحث إيجاد المدى الربيعي IQR باستخدام برنامج Spss عليه إتباع الأمر التالي.

⇨ Analyze
⇨ Descriptive Statistics
⇨ Frequencies
Click on the Statistics button
Check Range and Quartiles

المصدر J. Richard Kendrick Social Statistics: AN Introduction Using SPSS, USA, 2005, P. 214

إجراءات حساب مقاييس النزعة المركزية ومقاييس التشتت باستخدام Spss :

1- اختر A nalyze ⟸ Descriptive Statistics ⟸ Frequencies

2- اختر من قائمة المتغير (العمر Age).

3- انقر على ◄ يقوم بلصق Pastes العمر Age في القائمة المحددة للمتغيرات Variable(s).

4- انقر على Statistics يقودك إلى صندوق جديد بعنوان: Statistics : Frequencies

5- في المساحة المعنونة Central Tendency انقر على المربع إلى اليسار من Means، Median و Mode سيضع علامة ✓ على المربعات إشارة إلى تلك المقاييس التي نود حسابها.

6- في المساحة المعنونة Dispersion انقر على المربع إلى اليسار من Std. deviation و Range.

7- في المساحة المعنونة Percentile Values انقر على المربع إلى اليسار من Quartiles.

8- انقر على Continue

9- انقر على ok.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

Statistics

Age in years

N	Valid	20
	Missing	0
Mean		19.35
Median		19.00
Mode		18
Std. Deviation		1.35
Range		4
Perecntiles	25	18.00
	50	19.00
	75	20.00

شكل (4-4) مخرجات الإحصاء الوصفي باستخدام SPSS

أسئلة للمراجعة:

1- هل يمكنك حساب المتوسط الحسابي ($\overline{X}$) من البيانات الاسمية علل ذلك؟

2- ما هي ميزات وعيوب المدى كمقياس للتشتت؟

3- ماذا تعني الرموز التالية: ($\overline{X}$ ، σ، μ ,S)؟

4- إذا كان المتوسط الحسابي $\overline{X}$ لعدد ثمانِ درجات (8) هو 5 إذاً ما هو العدد المتبقي إذا كان عدد الدرجات: 7، 4، 6، 5، 4،3، 9.

5- من البيانات التالية: احسب كل من: المتوسط $\overline{X}$، والوسيط، والمدى، والانحراف المعياري من كل واحد من هذه التوزيعات.

أ- 5، 9، 13، 15، 26، 72.

ب- 121، 134، 145، 212، 289، 306، 367، 380، 453.

ج- 1.4، 1.2، 1.9، 2.0، 2.4، 3.5، 3.9، 4.3، 5.2.

6- من البيانات التالية:

أ- احسب المتوسط، والوسيط والمنوال.

ب- احسب المدى، والمدى الربيعي، والانحراف المعياري.

12	40	22	30	18	36	45	19
22	22	37	35	72	28	36	29
42	56	52	35	16	23	37	26
22	29	35	62				

7- من البيانات التالية:

F	X
127	4-1
500	8-5
784	12-9
59	16-13
8	20-17

أ- احسب المتوسط، والوسيط، والمنوال لهذا التوزيع.

ب- إذا كان ثمة فرقاً، اشرح لماذا؟

8- جد وفسر (IQV) من البيانات التالية:

درجة الرضا المهني	التكرار
راضٍ جداً	91
راضٍ	28
غير راضيٍ جداً	9
غير راضٍ	31
لا أستطيع أن أقرر	23
المجموع	**180**

9- أوجد، وفسر المقياس الملائم للنزعة المركزية و IQV للتوزيع التكراري التالي:

المستوى التعليمي	التكرار
جامعي فما فوق	81
تعليم متوسط	42
الشق الأول من التعليم الأساسي	40
الشق الثاني من التعليم الأساسي	35
لا تعليم	2
المجموع	**200**

الهوامش والمصادر:

أولا: الهوامش:

1- J. Richard Kendrick, Social Statistics: An Introduction using SPSS, Pearson Education INC, USA, 2005, P. 158.

2- Ibid, P. 158.

3- George Diekhoff, Statistics for Behavioral Sciences: Univariate Bivariate, Multivariate, The Mc Graw Hill Co. Inc, USA, 1992, P. 41.

4- Ibid, P. 43.

5- George Argyrous, Statistics for Social and Health Research With Guide to SPSS, Sage Publications, 2001, PP. 79 - 80.

6- Ibid, P. 80.

7- Ibid, PP. 80 - 81.

8- Ibid, PP. 82 - 83.

9- J. Richard Kendrick, op. Cit, P.186.

ثانيا: المصادر:

1- George Argyrous, Statistics for Social and Health Research Witha Guide to SPSS, Sage Publications, London, 2001 .

2- George Diekhoff, Statistics for Behavioral Sciences: Univariate Bivariate, Multivariate, The Mc Graw Hill Co. Inc, USA, 1992.

3- J. Richard Kendrick, Social Statistics: An Introduction using SPSS, Pearson Education INC, USA, 2005.

الفصل الخامس

المنحنى الطبيعي

مقدمة

في هذا الفصل سوف نتحدث بالتفصيل عن خصائص واحد من أكثر التوزيعات التكرارية الخاصة، ألا وهو التوزيع الطبيعي. ولا يعني مصطلح الطبيعي أنه معتاد أو مألوف، فالمنحنى الطبيعي لا يمكن النظر إليه في أي شيء إلا كونه طبيعياً. ومن هنا يلعب المنحنى الطبيعي دوراً رئيسياً في اختبار الفروض فهو الأساس لكثير من الإجراءات التي سنبينها في الفصول اللاحقة.

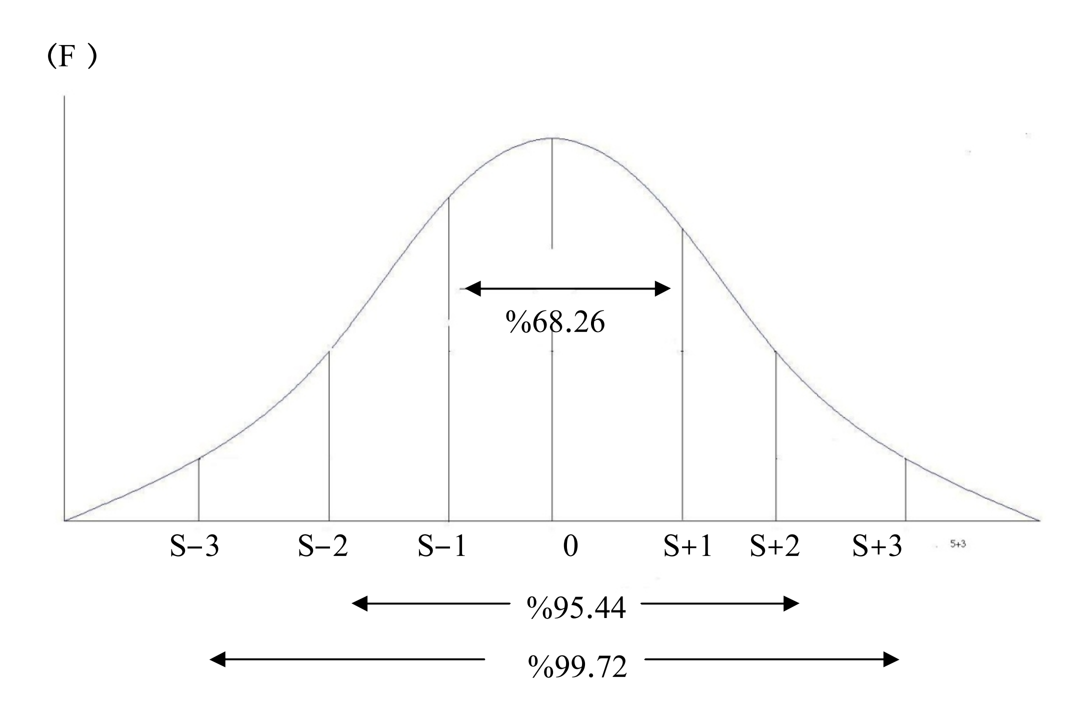
إن الأسباب الكامنة وراء دراسة المنحنى الطبيعي قد تكون في بادئ الأمر غير واضحة للقارئ، إلا أننا نأمل أن يكون الموضوع أكثر وضوحاً مع المواضع اللاحقة لاستخدامات المنحنى الطبيعي.

التوزيع الطبيعي:

التعريف: بداية دعنا نبدأ بتعريف بسيط للتوزيع الطبيعي آملين في التوسع في هذا التعريف عندما نكون على دراية كاملة بهذا التوزيع.

إن التوزيع الطبيعي هو نوع خاص من التوزيعات التكرارية للمجتمع. والمنحنى الطبيعي لهذا التوزيع يكون متماثلاً وله قيم متساوية للوسط الحسابي، والمنوال، والوسيط. وله معادلة رياضية خاصة به. وله وسط حسابي وانحراف، وله نسبة 68.3 من المساحة تحت المنحنى داخل انحراف معياري عن الوسط الحسابي. أي أنه ذلك التوزيع الطبيعي الذي وسطه صفراً (0) وتباينه يساوي (1)؛ وأن المساحة الكلية الواقعة تحت المنحنى تساوي 1.00.

إن هذه الخصائص المتعلقة بالمنحنى الطبيعي يمكن توضيحها في الشكل التالي:



شكل رقم (5-1): المساحات تحت المنحنى الطبيعي المعياري

يمكننا من خلال هذا الشكل أن نستنتج نتائج محددة حول التوزيع التكراري للدرجات التي نعتقد أنها موزعة توزيعاً طبيعياً [1].

على سبيل المثال [2] قد يكون الباحث راغباً في معرفة أعمار أفراد مجتمع ما ولدَيْه الإحصاءات الوصفية للوسط الحسابي. وقد توزعت هذه الأعمار لهذا المجتمع المكون من 400 فرداً على النحو التالي:

μ = 35 سنة

σ = 13 سنة

إذا كان العمر موزعاً توزيعاً طبيعياً فإن 68.3 % من أفراد هذا المجتمع ستقع داخل 1 انحراف معياري عن المتوسط. بمعنى آخر، أن 272 شخصاً (68.3 % من 400) سوف تكون أعمارهم في أي مكان بين 22 سنة (35 − 13) و48 سنة (35 + 13).

إن هذه الخاصية للمنحنى الطبيعي تبقى صحيحة بغض النظر عن أي قيمة محددة للانحراف المعياري والمتوسط لمجموعة الحالات التي نتعامل معها. فعلى سبيل المثال، يمكن أن يكون لدينا ثلاث مجموعات مختلفة لعدد 400 شخص بأعمار وصفت بالإحصاءات التالية (انظر جدول 5 − 1).

جدول رقم (5ـ1) متوسط العمر وتوزيعه على ثلاثة مجتمعات

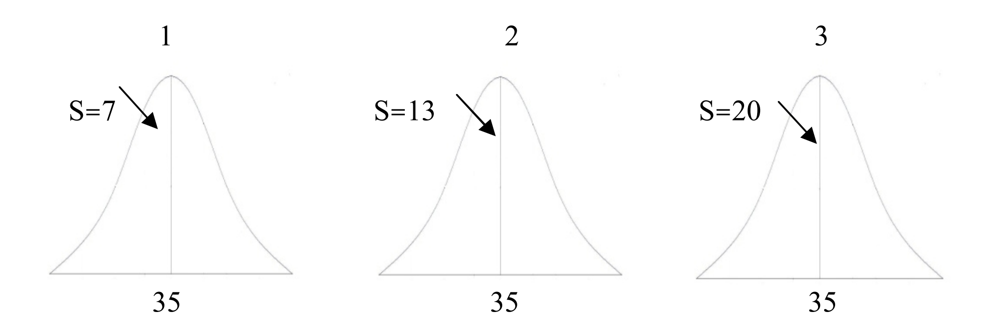
المجموعة	متوسط العمر (بالسنوات)	انحراف معياري (بالسنوات)	مدى متوسط العمر 68.3% للحالات
1	35	13	22 إلى 48
2	35	7	28 إلى 42
3	35	20	15 إلى 55

المصدر: George Argyrous, Statistics for Social and Healh Research, P. 113.

إن كل هذه التوزيعات الثلاثة لديها نفس المتوسط العمري، ولكنها تختلف من حيث توزيع الأعمار حول المتوسط. فإذا كان باستطاعتنا أن نفترض أن كل واحد من

هذه المجتمعات الثلاثة موزعة توزيعاً طبيعياً حينئذٍ يمكننا أن نستنتج متوسطات المدى العمري 68.3 % من الناس في كل مجموعة.

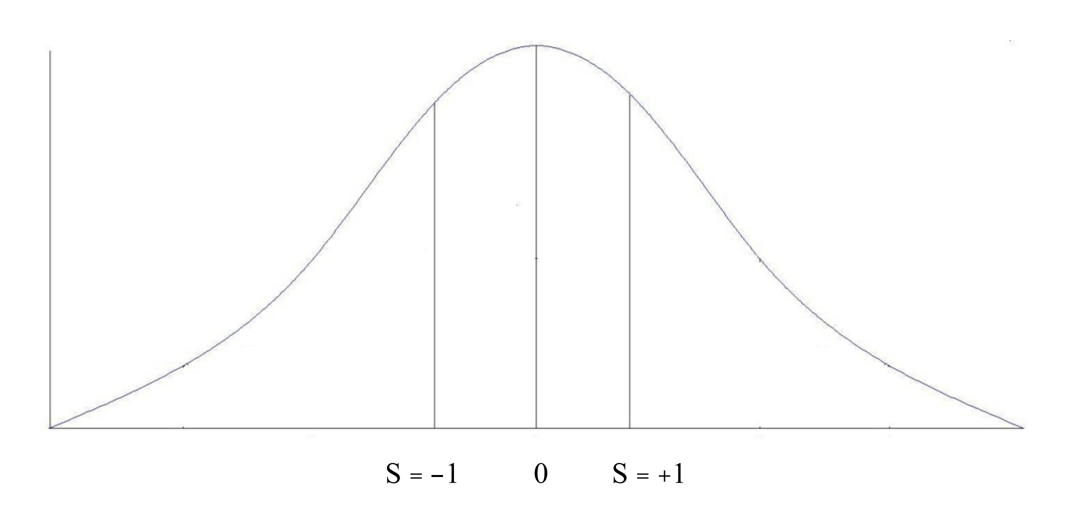
وإذا كانت هذه التوزيعات طبيعية، فإن 272 من الناس في مجموعة واحدة سوف تكون أعمارهم بين 22 – 48 سنة؛ و272 شخصاً في مجموعة 2 سوف تكون أعمارهم تتراوح من 28 إلى 42 سنة. والمجموعة الثالثة يكون مدى أعمارها من 15 إلى 55 سنة.



شكل (5-2) ثلاثة توزيعات طبيعية بانحرافات معيارية مختلفة

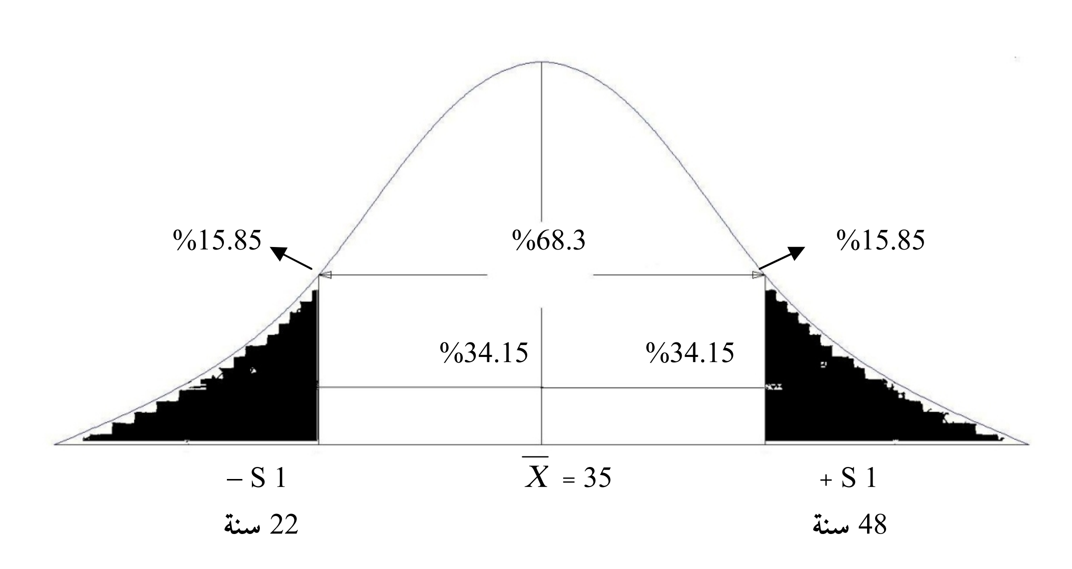
إن صياغة هذه العملية لتوزيع الحالات بالنظر إلى عدد الانحرافات المعيارية من المتوسط يطلق عليه (معايرة التوزيع) Standardizing the Distribution أي التوزيع الطبيعي المعياري. ونقصد هنا بالتوزيع الطبيعي المعياري، هو ذلك التوزيع الطبيعي الذي تم فيه تحويل الدرجات الإحصائية إلى درجات معيارية تعبر عن درجات أي توزيع تكراري طبيعي (انظر الشكل 5 – 3).





شكل (3-5) التوزيع الطبيعي المعياري

إن إجراء المعايِّرة يسمح لنا بقياس كل التوزيعات الطبيعية في إطار الوحدات الشائعة – وحدات الانحراف المعياري – بغض النظر عن الوحدات التي تم مبدئياً قياس المتغير عليها. وأن ما يمكن ملاحظته أن المنحنى الطبيعي هو توزيع متماثل باعتبار أن المنحنى متماثل فإن نفس نسبة الحالات التي تقع داخل مدى محدد فوق المتوسط الحسابي. كذلك تقع داخل نفس المدى تحت المتوسط الحسابي (انظر الشكل رقم 4). بمعنى آخر، إذا كانت 68.3 % لكل الحالات تقع داخل وحدة انحراف معياري واحد في أي من الاتجاهين من المتوسط فإن نصف هذه الحالات (34.15 %) سوف تقع فوق المتوسط الحسابي، وأن النصف الآخر (34.15 %) سوف يقع تحت المتوسط الحسابي لمجموعة (1) في الجدول رقم (1). سوف ينطوي على 136 شخصاً ستكون أعمارهم بين 22 و 35 سنة، والمجموعة الأخرى 136 شخصاً سيكون أعمارهم بين 35 و 48 شخصاً.

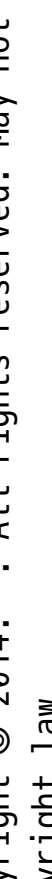


شكل رقم (5ـ4) توزيع العمر لمجموعة رقم (1)

إن الأمر الآخر، الذي يمكن ملاحظته حول توزيع الحالات تحت المنحنى الطبيعي في الشكل رقم (4) هو أن نسبة الحالات التي تقع أبعد من 1 انحراف معياري عن المتوسط الحسابي تكون مساوية للعدد الكلي للحالات 100 % ناقص النسبة التي تقع داخل المدى 68.3 %.

$$100.0 - 68.3 = 31.7\ \%$$

مرة أخرى، يمكننا، أن نقسم هذه المنطقة إلى منطقتين تكون نسبة 15.85 % من الحالات أعمارها فوق 1 انحراف معياري عن المتوسط (على سبيل المثال مجموعة (1) هذه المجموعة يكون أعمار أفرادها أكبر من 48 سنة)، والأخرى بنسبة 15.85 من الحالات على نهاية الجانب الآخر من المنحنى. وعليه، إذا كانت سيدة في هذه المجموعة قد أوضحت لنا أن عمرها 52 سنة، فإننا بالتالي نعرف أن هذه السيدة تقع في أكبر 16 % في المجتمع. إننا نأمل من خلال هذا التمرين البسيط الذي بيناه للتو، أن يكون مفيداً في فهمنا للمنحنى الطبيعي. وإذا كان بمقدورنا أن نفترض أن التوزيعَ التكراريَّ توزيعُ طبيعيٌّ،



حينئذٍ نستطيع أن نصل إلى نتيجة حول مدى القيم التي تقع خلالها نسبةٌ من الحالات المحددة عند المتوسط الحسابي والانحراف المعياري. المفترض أن مثل هذا الأمر يجعل استخدام المنحنى الطبيعي مهمّاً وذلك لسببين أساسيين:

المنحنى الطبيعي كأداة مساعدة لوصف البيانات:

هناك بعض التوزيعات الأمبيريقية في الواقع الحياتي على نحو أقرب أن تكون طبيعية، وتسمح لنا بتحديد نسبة معينة من الحالات التي تقع في مسافة محددة فوق و / أو تحت المتوسط. على سبيل المثال، قد نجد كثيراً من الخصائص الفيزيقية لدى الناس مثل الطول، تكون قريبة من التوزيع الطبيعي. فإذا أخذت عينة عشوائية من الأفراد وتم قياس الطول لهذه العينة، فإننا في واقع الأمر نجد 68 % من الحالات تقع داخل انحراف معياري واحد من المتوسط الحسابي.

المنحنى الطبيعي كأداة للإحصاء الاستدلالي:

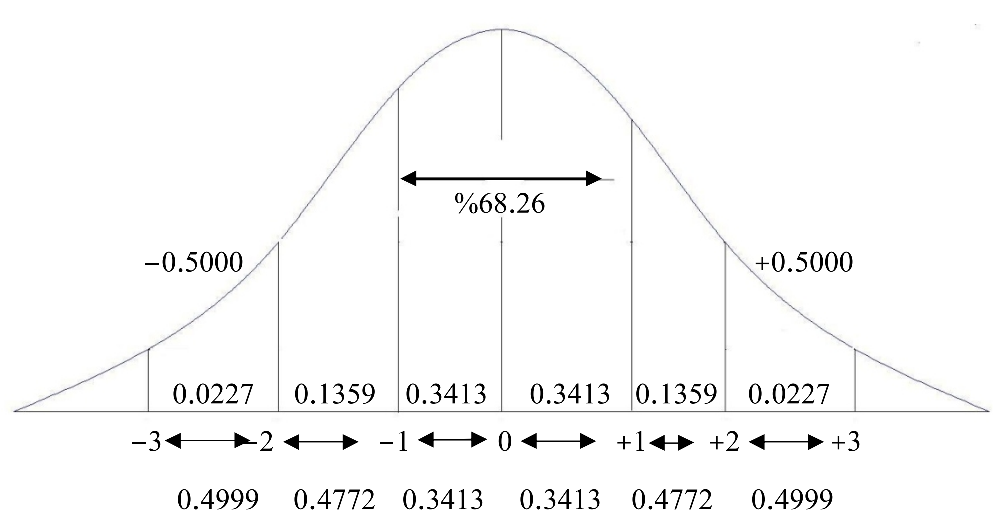
إن السبب الثاني لفهم خصائص المنحنى الطبيعي، والاحتمالية الأكثر أهمية تكون الأساس في تشكيل الإجراءات الأساسية التي تسمح لنا بإجراء الاستدلالات من العينة العشوائية على المجتمع (أي من الجزء إلى الكل).

إن أهمية المنحنى الطبيعي في الاستدلالات الإحصائية سوف يتم توضيحه من خلال بعض الجوانب التطبيقية في الجزء الثالث من هذا الكتاب.

استخدام المنحنيات الطبيعية لوصف توزيع:

تكمن المهمة الباقية لهذا الفصل في استخدام المنحنى الطبيعي كأداة للوصف مرجئين استخدامه كأداة للاستدلال من العينة على المجتمع للجزء الثالث من هذا الكتاب. أما الآن فيمكننا أن نتوسع قليلاً في تعريف المنحنى الطبيعي وذلك بتعريف نسبة المساحة الكلية تحت المنحنى الطبيعي داخل وحدتين، انحراف معياري من المتوسط الحسابي وداخل ثلاث وحدات انحراف معياري.

- بين ± 1 انحراف معياري عن المتوسط الحسابي لتوزيع طبيعي يغطي 68.2 % من المساحة تحت المنحنى.
- بين ± 2 انحراف معياري عن المتوسط الحسابي لتوزيع طبيعي يغطي 95.44 % من المساحة تحت المنحنى.
- بين ± 3 انحراف معياري عن المتوسط الحسابي لتوزيع طبيعي يغطي نسبة 99.72 % من المساحة تحت المنحنى.



هذه المعلومات يمكننا توضيحها في الجدول البسيط التالي (جدول 5 – 2).

جدول (5-2) المساحات تحت المنحنى الطبيعي المعياري

الانحراف المعياري انحرافات معيارية عن المتوسط	المساحة تحت المنحنى بين كلا النقطتين	المساحة تحت المنحنى أبعد من كلا النقطتين (ثنائي الجانب)	المساحة تحت المنحنى أبعد من نقطة واحدة (أحادي الجانب)
1 ±	0.683	0.317	0.1585
2 ±	0.954	0.046	0.0230
3 ±	0.996	0.004	0.0020

هناك مظهران أساسيان تجدر الإشارة إليهما في جدول (5 – 2).

- بدلاً من التعبير عن المساحة تحت المنحنى الطبيعي كنسبة، ثم التعبير عنها كتناسب: 68.3 % ثم تحويلها إلى 0.683 وهكذا....
- إن القيم الموجودة في العمودين الأولين سوف يصل دائماً مجموعهما إلى واحد (على سبيل المثال 0.683 + 0.317 = 1) وهذا يرجع إلى أن المساحتين معاً يجب أن تكونا مساويتين للمساحة الكلية تحت المنحنى الذي يصل إلى 100 %.

إن المنحنى الطبيعي يمكن تعريفه بشكل محدد كمضلع أو نمط من الرسم يسمح لنا أن نفسر التناسب في الجدول كاحتماليات. فالاحتمالية في هذا السياق هي ببساطة الفرصة لأي حالة معطاة سوف يكون لديها قيمة محددة، أو تقع داخل مدى محدد من القيم. فعلى سبيل المثال، نفترض أن شخصاً ما تم اختياره بشكل عشوائي من مجموعة (1) وعلينا أن نخمّن كم يكون عمر هذا الشخص. في هذه الحالة يمكننا استخدام جدول

المساحة تحت المنحنى الطبيعي لنقرر احتمالية عمر هذا الشخص تكون في مكان ما بين 22 سنة و 48 سنة (على سبيل المثال، يكون عمر هذا الشخص داخل واحد انحراف معياري أي من جانبي المتوسط). 0.683 أو حول 68 %. إن احتمالية أنّ هذا الشخص يصل عمره إلى أقل من 22 سنة تكون 0.1585 أو حول 16 %. منطقياً عادة ما توجد احتمالية عالية أن شخصاً ما تم اختياره عشوائياً من مجموعة من الحالات سوف يعكس المتوسط الحسابي لهذه الحالات أنه على الأرجح أن شخصاً ما سوف يكون نموذجاً بدلاً من كونه استثنائياً، هذه هي طريقة تفسير المساحة تحت المنحنى الطبيعي، كاحتمالية ستكون بشكل خاص ذات أهمية في الفصول اللاحقة المتعلقة بالاستدلال الإحصائي [3].

جدول رقم (5-3) المساحات تحت المنحنى الطبيعي المعياري

المساحة تحت المنحنى أبعد من نقطة واحدة	المساحة تحت المنحى أبعد من كلتا النقطتين	المساحة تحت المنحنى بين كلتا النقطتين	Z
0.4600	0.920	0.080	± 0.1
0.4205	0.841	0.159	± 0.2
0.3820	0.764	0.236	± 0.3
0.3445	0.689	0.311	± 0.4
0.3085	0.617	0.383	± 0.5
0.2745	0.549	0.451	± 0.6
0.2420	0.484	0.516	± 0.7
0.2120	0.424	0.576	± 0.8
0.1840	0.368	0.632	± 0.9
0.1585	0.317	0.683	± 0.1
0.1355	0.271	0.729	± 1.1
0.1150	0.230	0.770	± 1.2
0.0970	0.194	0.806	± 1.3

Z	المساحة تحت المنحنى بين كلتا النقطتين	المساحة تحت المنحى أبعد من كلتا النقطتين	المساحة تحت المنحنى أبعد من نقطة واحدة
± 1.4	0.838	0.162	0.0810
± 1.5	0.866	0.134	0.0670
± 1.6	0.890	0.110	0.0550
± 1.645	0.900	0.100	0.0500
± 1.7	0.911	0.089	0.0445
± 1.8	0.928	0.072	0.0360
± 1.9	0.943	0.057	0.0290
± 1.96	0.950	0.050	0.0250
± 2	0.954	0.046	0.0230
± 2.1	0.964	0.036	0.0180
± 2.2	0.972	0.028	0.0140
± 2.3	0.979	0.021	0.0105
± 2.33	0.980	0.020	0.0100
± 2.4	0.984	0.016	0.0080
± 2.5	0.988	0.012	0.0060
± 2.58	0.990	0.010	0.0050
± 2.6	0.991	0.009	0.0045
± 2.7	0.993	0.007	0.0035
± 2.8	0.995	0.005	0.0025
± 2.9	0.996	0.004	0.0020
± 3	> 0.996	< 0.004	< 0.0020

درجات z:

إن الدرجة المعيارية لـِ z توضح كم انحرافاً معيارياً تبتعد الدرجة الأصلية عن الوسط الحسابي؟ وبدلاً من استخدام تعبير عدد الانحرافات المعيارية عن المتوسط سوف تستخدم درجات z فدرجة +1 لـِ z تشير إلى واحد انحراف معياري فوق المتوسط. ودرجة -1.5 لـِ z تشير إلى 1.5 انحرافات معيارية تحت المتوسط. ولمجتمع طبيعي أو عينة طبيعية، يمكننا التعامل مع درجة z مرتبطة مع أي قيمة فعلية مستخدمين المعادلة التالية:

$$Z = \frac{X - \mu}{\sigma}$$

$$Z = \frac{X - \overline{X}}{S}$$

حيث إن: X هي القيمة الفعلية تم قياسها على الوحدات الأصلية.

- μ هي متوسط المجتمع.
- σ هي الانحراف المعياري للمجتمع.
- $\bar{x}$ متوسط العينة.
- S الانحراف المعياري للعينة.

يمكننا على سبيل المثال دراسة مجتمع مؤلف من 400 شخص في مجموعة 1 أعلاه، بمتوسط عمري 35 سنة وانحراف معياري 13 سنة. وإن أحد أعضاء هذه المجموعة أوضح لنا أنه يبلغ من العمر 61 سنة، حتى قبل أن نُعّقد المسألة من خلال المعادلة ودرجات Z فإنه من الوضوح أنّ هذا الشخص قد يكون في عمر أكبر من المتوسط الحسابي، ولذلك بداهةً يمكننا أن نستنتج أن قلةً قليلةً جداً من الناس سوف تكون في هذا العمر أو أكبر. في الحقيقة، إنه باستطاعتنا حتى هذه النقطة أن نستخدم الوصف التعبيري ونقول إن المتوسط والانحراف المعياري المحدد لهذه المجموعة فقط قلة قليلة من الناس سوف تصل أعمارها إلى 61 سنة أو أكثر. وباستطاعتنا على أية حال، أن نكون أكثر دقة من هذا وأن نحسب كم نسبة هذه المجموعة الصغيرة للمجموع الكلي؟ ولفعل ذلك يمكننا وضع

المعلومات في المعادلة لحساب درجات Z للمجتمع وحساب درجة Z المرتبطة بهذا العمر:

$$Z = \frac{X-\mu}{\sigma} = \frac{61-35}{13}$$

$$= 2$$

هذه النتيجة مباشرة تبين لنا أنّ 61 تكون 2 انحرافات معيارية فوق المتوسط وبالإشارة إلى العمود الأخير في جدول رقم (2) المساحات تحت المنحنى الطبيعي يوضح أنّ نسبة الناس الذين تبلغ أعمارهم 61 سنة أو أكثر هم فقط 0.023 أو 2.3 % من المجموع الكلي.

وفي الحقيقة، لقد أنجز الإحصائيون جداول إحصائية تتضمن قيم المساحة تحت المنحنى الطبيعي بين المتوسط وكل نقطة على طول المحور الأفقي للمنحنى الطبيعي، كما نلاحظ ذلك في جدول المساحات تحت المنحنى الطبيعي (جدول رقم 3).

إنه من الضرورة بمكان حتى هذه النقطة أن نكرر لماذا نشغل أنفسنا بتعريف المنحنى الطبيعي بهذه التفاصيل الدقيقة؟ كذلك يمكننا طرح السؤال التالي: لماذا يذهب الإحصائيون إلى مثل هذه التفاصيل كعمل فعلي ثم طباعة الجداول التي تشير إلى عدد الحالات التي تقع داخل مناطق محددة للتوزيع الطبيعي؟ بعد كل هذا يوجد لدينا عدد لا متناهي من التوزيعات التكرارية المحتملة التي تواجهنا - التوزيع العائلي وفقاً للدخل الكلي سوف يختلف عن توزيع المدن طبقاً لمعدلات الجريمة - وعليه يمكننا القول، إن كلا هذين التوزيعين لن يشبها إلى حد بعيد التوزيع الطبيعي. فالسؤال المطروح هنا هو: لماذا لا نقوم ببناء جداول تحدد لنا المساحات تحت المنحنيات [4].

أولاً: توجد مجموعة من التوزيعات الأمبيريقية التي تقرب من التوزيع الطبيعي، وعليه فإن مثل هذا الجدول سوف يكون مفيداً في مساعدتنا في وصف مثل هذه التوزيعات.

ثانياً: ولما كان هناك وجود توزيع طبيعي مرتبط بشكل أساسي بالإحصاءات الاستدلالية سنبيّن ذلك في الفصول اللاحقة التي تعالج المنحنى الطبيعي على نحو استثنائي مفيد في عملية الاستدلال من عينة على المجتمع. ومن خلال الفصل سنورد مجموعة من

الأمثلة لتساعدنا على فهم أفضل حول الكيفية التي يمكن من خلالها استخدام المنحنى الطبيعي المعياري كأداة للوصف. ومن خلال هذه العملية سنعمل أيضاً على تعويد أنفسنا بالإجراءات للنظر إلى جدول القيم في المساحة تحت المنحنى الطبيعي المعياري، والتي سوف تكون ذات أهمية بالغة في الفصول اللاحقة من هذا الكتاب.

على سبيل المثال، نفترض أن لدينا امتحان من 100 درجة لعينة مؤلفة من (100) مائة طالب. وتحصلنا على النتائج الإحصائية التالية:

$$\overline{X} = 60$$

$$S = 10$$

وإذا رسمنا هذه البيانات على مضلع تكراري فإننا نشاهد أن هذا التوزيع يشبه تقريباً التوزيع الطبيعي. يوجد عدد من المقاييس الإحصائية لتقييم المدى الذي يكون فيه التوزيع توزيعاً طبيعياً وسنبين هذه الطرق عند الحديث عن تقدير فترات الثقة. إننا هنا في هذا المثال، قد اعتمدنا على الفحص العيني للمضلع التكراري لنقرر ما إذا كان التوزيع يقرب من الطبيعي. لمعرفة ذلك. فإن مجموعة هؤلاء الطلاب قد توزعت توزيعاً طبيعياً (أو قريباً منه) طبقاً لدرجات الامتحان لتسمح لنا بالإجابة على مجموعة عديدة من الأسئلة حول هذا المتغير.

المساحة بين المتوسط الحسابي ونقطة على التوزيع:

ربما يريد الباحث معرفة كم عدد الطلاب بين متوسط 60 و 65 درجة التي يمكن أن يعتقد بأنها تشكل مدى معقول من الدرجات يمكن للطلاب أن ينجزوها.

إن أول شيء يمكن فعله هو تحويل درجة 65 إلى درجة Z المعيارية.

$$Z = \frac{X-\mu}{\sigma} = \frac{65-60}{10}$$

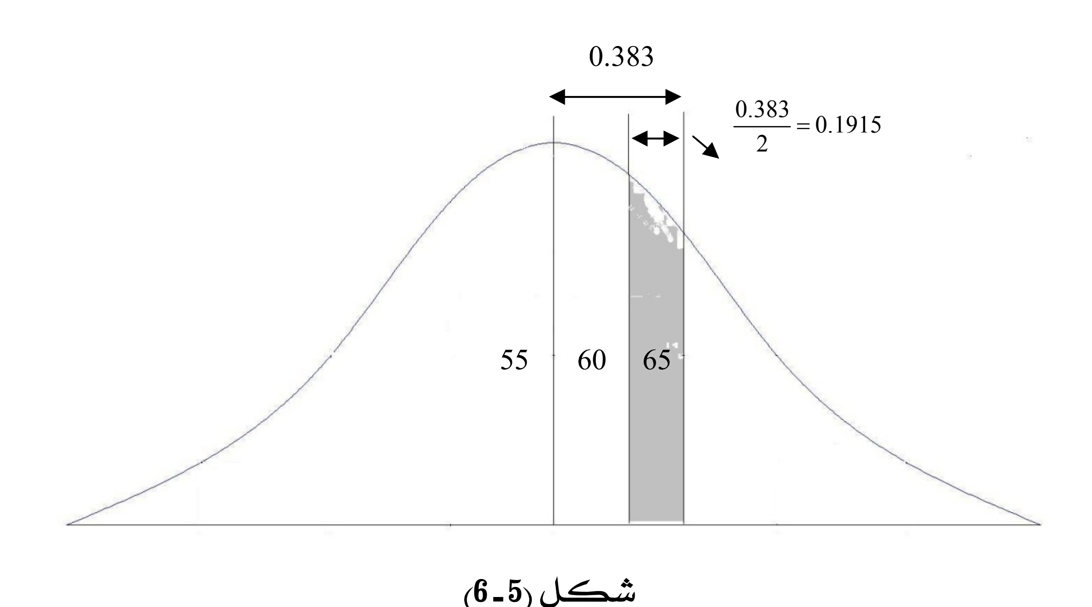
$$= 0.5$$

إن الخطوة التالية هي الرجوع إلى جدول المساحات تحت المنحنى الطبيعي المعياري لنجد المساحة بين هذه النقطة والمتوسط. ومن خلال النظر إلى هذه الجزئية من المساحات تحت المنحنى الطبيعي المعياري لدرجة Z=0.5 ستتحصل على النتيجة المبينة في الجدول رقم (3).

بمعنى آخر، أن 0.383 لكل الحالات سوف نتحصل على 5 درجات فوق أو تحت المتوسط. ولما كُنّا نرغب فقط في أولئك الطلاب الذين سجلوا 5 درجات فوق المتوسط الحسابي، فإنه يمكننا أن نقسم 0.383 إلى نصفين. نسبة الطلاب الذين سجلوا درجات بين 60 و 65 $\frac{0.383}{2}$ تساوي 0.1915 (انظر الشكل 5 – 6 لزيادة التوضيح).

جدول (5-4) المساحات تحت المنحنى الطبيعي المعياري

Z	المساحة تحت المنحنى بين كلتا النقطتين	المساحة تحت المنحى أبعد من كلتا النقطتين	المساحة تحت المنحنى أبعد من نقطة واحدة
± 0.1	0.080	0.920	0.4600
± 0.2	0.159	0.841	0.4205
± 0.3	0.236	0.764	0.3820
± 0.4	0.311	0.689	0.3445
± 0.5	0.383	0.617	0.3085
± 0.6	0.451	0.549	0.2745
± 0.7	0.510	0.484	0.2420
± 0.8	0.576	0.424	0.2120
± 0.9	0.632	0.368	0.1840
± 1.0	0.683	0.317	0.1585
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮
± 3	> 0.996	< 0.004	< 0.0020



شكل (5-6)

وعليه يمكننا القول إنه فقط أكثر من 0.19 (19%) من الطلاب تحصلوا على درجات بين 60 و 65 (لاحظ أنه يمكن تحويل التناسب إلى نسبة بتحريك الفاصلة العشرية مَوّضِعَيْن إلى اليمين).

المساحة أبعد من نقطة على التوزيع:

تجدر الإشارة إلى أن نفس المنطق الذي تم استخدامه في المثال السابق يمكن تطبيقه هنا لإيجاد نسبة الحالات التي تقع أبعد من نقطة محددة على التوزيع. فعلى سبيل المثال، قد نرغب في معرفة نسبة الطلاب المميَّزين الذين سجلوا درجات تفوق 65 درجة. ومن المثال السابق يمكننا معرفة درجة Z المرتبطة بدرجة 65.

$$Z = \frac{X - \overline{X}}{S} = \frac{65 - 60}{10}$$

$$= 0.5$$

في هذه المرة عند الرجوع إلى جدول المنحنى الطبيعي المعياري فإننا ننظر إلى عمود المساحة أبعد من نقطة ثم تحديدها بدرجة $Z = 0.5$.

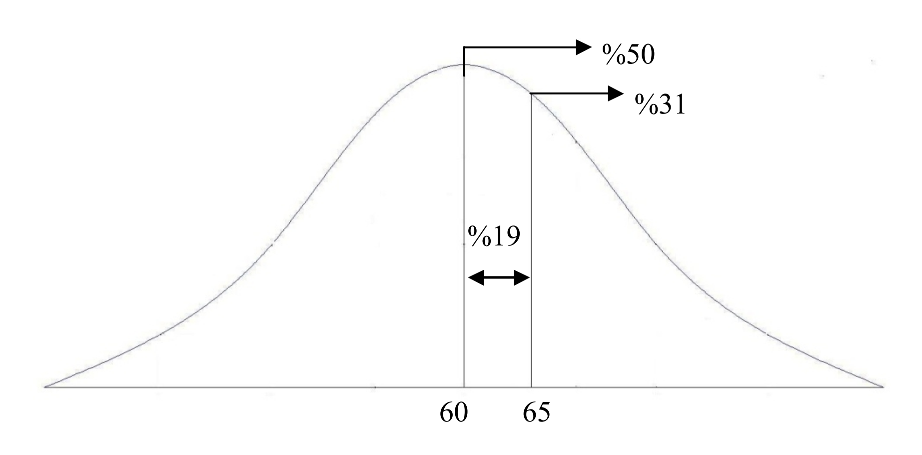
بمعنى آخر، أننا نرغب فقط في المساحة تحت جانب واحد من التوزيع (انظر جدول المساحات تحت المنحنى الطبيعي المعياري).

يشير هذا الجدول إلى أن 0.3085 (30.85%) من الطلاب تحصلوا على 65 درجة فما فوق. وإذا ما نظرنا إلى الإجابة لهذين المثالين فإنه باستطاعتنا أن نرى أن مجموع نسبة هاتين الحالتين تساوي 50 (انظر الجدول رقم 5 – 5). المساحات تحت المنحنى الطبيعي.

ويرجع السبب في ذلك إلى أن المساحتين اللّتَيّن تم تحديدهما معاً تعطيان بالضبط نصف مساحة المنحنى (شكل 5 – 7).

جدول رقم (5-5) المساحات تحت المنحنى

مدى درجات الامتحان	نسبة الحالات (%)
– بين 60 و 65	19 %
– 65 فما فوق	31 %
مج	50



شكل (5-7) المساحات تحت المنحنى الطبيعي

وبنفس الطريقة قد نرغب في معرفة نسبة الطلاب الذين لم يحالفهم النجاح في الامتحان، وبالتالي يمكننا حساب درجة z المرتبطة بالدرجة التي تقل عن 50.

$$Z = \frac{X - \overline{X}}{S} = \frac{50 - 60}{10}$$

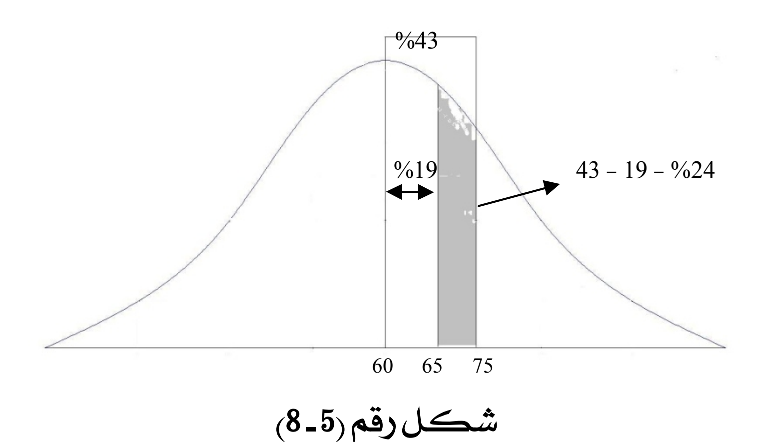
$$= -1$$

وبالرجوع إلى جدول المساحات تحت المنحنى الطبيعي المعياري نتحصل على 0.1586 من المنحنى أبعد من درجة z المساوية 1− والتي تشير إلى 16 % من الطلاب قد رسبوا في هذا الامتحان.

المساحة بين نقطتين على التوزيع الطبيعي:

سؤال آخر قد نرغب فيه، هو معرفة نسبة الحالات التي تقع داخل مدى ليس على حدود جانب واحد من المتوسط. على سبيل المثال، قد نرغب في معرفة نسبة الطلاب الذين تحصلوا على درجة معتمدة في سجلهم الدراسي والتي هي بين 65 و 75 درجة.

إن الإجابة على هذا السؤال المحير تكون واضحة عند النظر إلى الشكل رقم 8 . إن المساحة بين 65 و 75 درجة هي المساحة الباقية إذا ما طرحنا المساحة بين 65 والمتوسط من المساحة بين 75 والمتوسط. بمعنى آخر، فإننا نحتاج إلى حساب نسبتين تُحِطْنَ بالمتوسط و 65 وتلك التي تحيط بالمتوسط و 75.



شكل رقم (5ـ8)

من خلال المثال السابق ندرك أن 19 % من الحالات ستتحصل على درجات بين 60 و 65. ولتحديد نسبة الحالات التي سوف نتحصل على درجات بين 60 و 75 يتطلب منا أولاً حساب درجة z لهذا المدى من الدرجات:

$$Z = \frac{X - \overline{X}}{S} = \frac{75 - 60}{10}$$

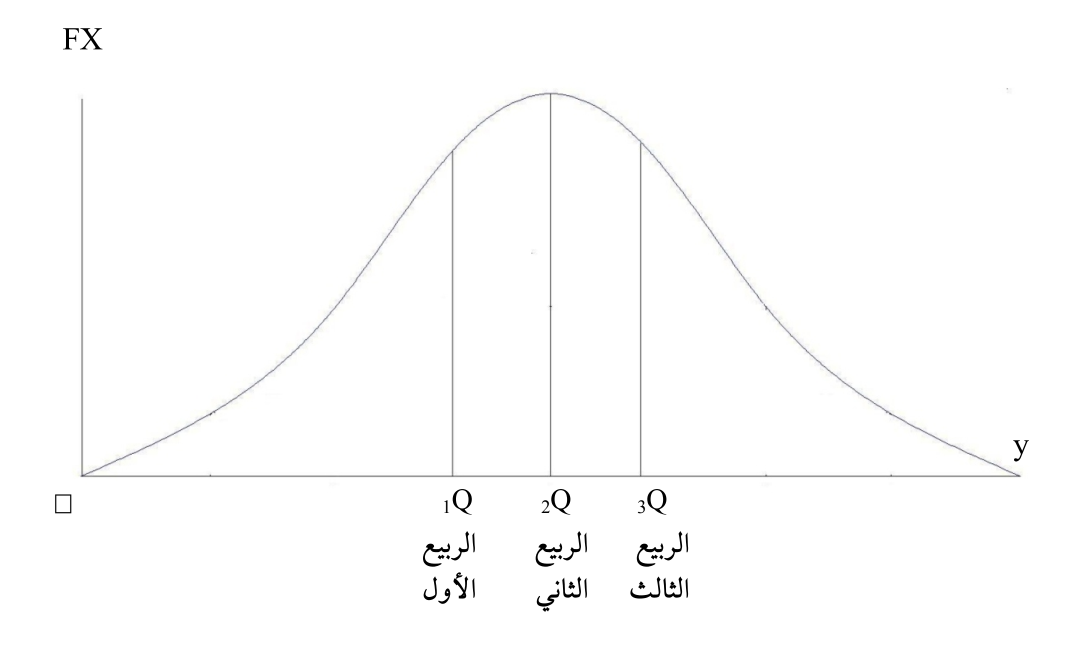
$$= 1.5$$

ومن خلال جدول المساحة تحت المنحنى الطبيعي المعياري نجد أن 0.866 (86.6%) من الحالات التي تقع على درجة z المساوية 1.5 أعلى وتحت المتوسط، لذلك فإن نصف هذه الحالات (43.3 %) سوف تقع درجاتها فوق المتوسط بدرجات بين 60 و 75. إن النتيجة هي 24 % من الطلاب قد تحصلوا على درجة معتمدة في سجلهم الدراسي.

حساب القيم من درجات z:

من خلال الأمثلة السابقة أردنا أن نبين نسبة الحالات التي لديها مدى محدد من الدرجات. وعلى أية حال فإن المشكلة التي نود طرحها هنا قد تكون مختلفة قليلاً نوعاً ما. وربما تكون قد حددت سلفاً نسبة الحالات التي كنا نرغب فيها وكنا نريد أن نستنتج مدى الدرجة التي تقع داخلها تلك النسبة. فعلى سبيل المثال، قد نرغب في معرفة مدى الدرجات التي تبين وسط 50 % من الطلاب. وبطريقة أخرى يمكننا طرح هذه الإشكالية بسؤال أي من هذه الدرجات تسجل الأعلى أو الأدنى المحاط بالمدى الربيعي.

تستخدم الرُّبَيْعَّات عادة في الامتحانات لمعرفة أي من الطلاب قد حصل على أقل من نسبة معينة وحصل باقي الطلاب على أكثر منه. بمعنى آخر، هي القيمة عند حد المئيني للتوزيع التكراري مقسماً إلى أربعة أجزاء كل جزء يحتوي على ربع من المجتمع. في حين نعني بالمئيني هو النسبة المئوية لمجموع التكرارات للقيم التي أقل من تلك القيمة بالنسبة لمجموع التكرارات الكلي كما يصفه الشكل التالي:



بالرجوع إلى جدول رقم (5 – 7) لإيجاد درجات z التي تسجل أو تبين منطقة 0.5 (50 %).

بالنظر إلى عمود المساحة تحت المنحنى بين نقطتين وإيجاد الخلية التي تكون احتمالاتها 0.5 أو الأقرب لها.

جدول (5-7) المساحات تحت المنحنى الطبيعي المعياري

المساحة تحت المنحنى أبعد من نقطة واحدة	المساحة تحت المنحى أبعد من كلتا النقطتين	المساحة تحت المنحنى الطبيعي بين النقطتين	درجة Z
0.4600	0.920	0.080	± 0.1
0.4205	0.841	0.159	± 0.2
0.3825	0.764	0.236	± 0.3
0.3445	0.689	0.311	± 0.4
0.3085	0.617	0.383	± 0.5
0.2745	0.549	0.451	± 0.6
0.2420	0.484	0.516	± 0.7
0.2120	0.424	0.576	± 0.8
0.1840	0.368	0.632	± 0.9
0.1585	0.317	0.683	± 1.0
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮
< 0.0020	< 0.004	> 0.996	± 3

إن أقرب قيمة إلى 0.5 من 0.516 التي ترتبط بدرجة z تساوي +0.7 و −0.7 ولتحويل درجات z +0.7 و −0.7 إلى وحدات فعلية (درجات الامتحان) التي على ضوئها تم قياس هذا المتغير، فالأمر يحتم علينا إعادة ترتيب المعادلة بشكل طفيف.

$$Z = \frac{x - \bar{x}}{s} \rightarrow x_1 = \bar{x} \pm z(s)$$

وإذا وضعنا درجة Z التي تحدد المنطقة في المعادلة سنتحصل على:

$$x_1 = \bar{x} \pm z(s)$$

$$x_1 = 60 + .07(10) = 67$$

$$x_1 = 60 - .07(10) = 53$$

وعليه فإن "الوسط" 50% من الطلاب قد سجلوا درجات بين 53 و 67 في الامتحان. وهذا يعني أيضاً أن 25 % من الطلاب هم تحت 53 و 25 % من الطلاب درجاتهم فوق 67 [5].

أسئلة للمراجعة:

1- عرف المنحنى الطبيعي المعياري وبين خصائصه؟

2- درجات امتحان كانت موزعة توزيعاً طبيعياً بمتوسط حسابي 10 وانحراف معياري 3. جد درجة Z لكل درجة من الدرجات التالية ونسبة المساحة فوق وتحت الدرجة؟

نسبة المساحة تحت	نسبة المساحة فوق %	درجة Z	X
			5
			6
			7
			8
			9
			11
			12
			14
			15
			16
			18

3- إذا كانت الدرجة التي تحصلت عليها في مادة الإحصاء الاجتماعي 78، وكان متوسط درجات الفصل في امتحان هذه المادة 67، بانحراف معياري 5. كيف تقارن درجتك بالتوزيع لكل درجات الاختبار؟

- أوجد درجة Z، ونسبة المساحة فوق وتحت الدرجة.

4- أُجْرِيَ امتحانُ نهائيٌّ في مادة تصميم البحوث الاجتماعية لكل المجموعات المسجلة في هذه المادة بجامعة قاريونس. ووزعت درجات الامتحان توزيعاً طبيعياً بمتوسط حسابي 72 درجة، بانحراف معياري 8. ما هي نسبة الطلاب الذين تحصلوا على درجة بين 60 و 69. وما هي نسبة الدرجة بين 70 و 79.

• المطلوب: حساب درجات Z، وبعد ذلك إيجاد المساحات بين كل درجة ومتوسطها الحسابي، وبعد ذلك اطرح المساحة الصغيرة من المساحة الكبيرة.

5- أُجْرِيَ امتحانُ شاملُ لطلاب الدكتوراة لتقييم قدراتهم في الترشح لدرجة الدكتوراة، وكان متوسط درجات هذا الامتحان الشامل 74، بانحراف معياري 10. ما هي نسبة درجات الطلاب المتحصلين عليها.

	درجة Z	المساحة
• بين 75 و 85		
• بين 80 و 85		
• فوق 80		
• فوق 83		
• بين 80 و 70		
• بين 75 و 70		
• تحت 75		
• تحت 77		
• تحت 80		
• تحت 85		

6- لقد تم إجراء دراسة لقياس التحامل الاجتماعي لعينة كبيرة من المبحوثين وقد وزعت الدرجات المتحصل عليها توزيعاً يقرب من الطبيعي، بمتوسط حسابي 31، وانحراف معياري 5. ما هي نسبة الدرجات المتحصل عليها في العينة؟

	درجة Z	المساحة
• تحت 20		
• تحت 40		
• بين 30 و 40		
• بين 35 و 45		

- فوق 25
- فوق 35

7- من جدول المساحة تحت المنحنى الطبيعي المعياري، أوجد الاحتمالية أن المتغير قد وزع توزيعاً طبيعياً ولديه درجة Z:

- فوق 1.3
- تحت 1.3
- بين 0.5 و 3.4
- بين 2.3– و 2
- أكبر من 2.3 وأقل من 1.4–
- أقل من 1.6– وأكبر من 1.96
- أقل من 1.96– وأكبر من 1.96

8- إذا كان متغير X موزعاً توزيعاً طبيعياً بمتوسط 60، وانحراف معياري 10. كم انحرافات معيارية من المتوسط الحسابي لقيم X التالية:

- 60
- 52
- 85
- 43
- 73

الهوامش والمصادر:

أولا: الهوامش:

1- Joseph F. Healey, The Essentials of Statistics: A Tool for Social Research, Wadsworth Cengage Learning, USA, 2010. pp. 127– 129.

2- George Argyrous, Statistics for Social and Health Research With a Guide to SPSS, Sage Publications, London, 2001, pp. 113 - 115.

3- 3. Ibid, PP. 115 - 118.

4- 4. Ibid, P. 120.

5- 4. Ibid, PP. 118 - 125.

ثانيا: المصادر:

1- Frederick J. Gravetter and Larry B. Wallnau, Statistics for The Behavioral Sciences, 8th ed., Wadsworth Cengage Learning, USA, 2010.

2- George Argyrous, Statistics for Social and Health Research Witha Guide to SPSS, Sage Publications, London, 2001.

3- Joseph F. Healey, The Essentials of Statistics: A Tool for Social Research, Wadsworth Cengage Learning, USA, 2010.

الجزء الثاني

الإحصاءات الوصفية الثنائية

- الفصل السادس: تحليل جداول التقاطع: استقصاء العلاقة بين متغيرين: الوصف الجدولي لبيانات مقاسة على المستويين الاسمي والترتيبي
- الفصل السابع: الوصف العددي للبيانات الاسمية: مقاييس التطابق الثنائية
- الفصل الثامن: التطابق بين متغيرات تم قياسها على المستوى الترتيبي
- الفصل التاسع: التطابق بين متغيرات تم قياسها على المستوى ذي المسافات والنسبي: شكل الانتشار وخط الانحدار

الفصل السادس

تحليل جداول التقاطع: استقصاء العلاقة بين متغيرين: الوصف الجدولي لبيانات مقاسة على المستويين الاسمي والترتيبي

مقدمة

في الفصول المتعلقة بالجزء الأول من هذا الكتاب تناولنا طرق التحليل المتعلقة بمتغير واحد في وقت محدد، وتُعرف هذه الطريقة بطريقة الإحصاءات الوصفية الأحادية. ويمكن أن تكون لدينا بيانات متعلقة بمتغير واحد، مثل الدخل، نستطيع وصفه بطريقة ملائمة تسمح لنا بقياسه طبقاً لأهداف البحث ومستوى قياس المتغير، فعلى سبيل المثال؛ باستطاعتنا وصف الدخل الأسبوعي مقاساً بالدينار الليبي وذلك من خلال الطرق الثلاثة التالية:

1- الوصف الجدولي: وذلك من خلال توليد جدول تكراري نسبي.

2- الوصف البياني: وذلك من خلال توليد مضلع تكراري.

3- الوصف العددي أو الرقمي وذلك من خلال إجراء العملية الحسابية للحصول على المتوسط الحسابي والانحراف المعياري لضمان النزعة المركزية والتشتت لهذا التوزيع؟

إن كل واحد من هذه الإجراءات يركز تحديداً على متغير واحد نرغب في دراسته.

العلاقة بين المتغيرات:

بطبيعة الحال، في خضم عملية إجراء البحث في العادة نقوم بجمع البيانات حول عدد كبير من المتغيرات، كالدخل، والتعليم، والنوع، والوضع الصحي، وعدد ساعات مشاهدة الإذاعة المرئية، وعدد ضخم من المتغيرات الأخرى التي يرغب الباحث في دراستها. وكل واحد منا يمتلك الإحساس العام، للفكرة التي مفادها أن متغيرين قد يرتبطان بعضهما ببعض، أو أنهما يعتمدان على بعضهما البعض. ونحن نعرف أنه كلما كبر الأطفال زادوا طولاً، فالعمر والطول متغيران يرتبطان ببعضهما البعض. كما أننا أيضاً نعلم أنه كلما زاد دخلنا زاد نمط استهلاكنا، فالدخل والاستهلاك يرتبطان ارتباطاً وثيقاً. إن هذه الأمثلة توضح لنا المفهوم العام الذي نشعر به. إن هذه الفكرة تبين لنا أنه كلما تغيرَّت إحدى قيم المتغير من حالة إلى حالة أخرى تغيرِّت معها أيضاً قيم المتغير الآخر، عادة يختار الباحث متغيرين لاعتقاده أنهما يرتبطان بعضهما ببعض بطريقة متسقة. فعلى سبيل المثال، قد يُعْتَقَدُ أن دخل شخص يرتبط بشكل ما بالمكان الذي يقطن فيه. ولبحث هذه العلاقة يتطلب ذلك جمع بيانات من عينة من الناس، قد نجد من خلال دراسة هذه العينة أن الناس الذين يعيشون في أحدى المدن لديهم مستوى اقتصادي محدد، في حين أن الناس الذين يعيشون في مدينة مختلفة يميلون إلى مستوى اقتصادي مرتفع، بينما الناس الذين يعيشون في مدينة ثالثة يسجلون مستوى اقتصادي أكثر ارتفاعاً. إن مثل هذه النتائج تمكننا من استنتاج أن مكان الإقامة ومستوى الدخل يرتبطان حقاً ارتباطاً وثيقاً. فإذا كانت العلاقة بين الإقامة ومستوى الدخل علاقة ارتباطية، يعني هذا أننا نتوقع أن الأشخاص يختلفون في دخولهم وفقاً لمكان الإقامة. ومن هنا يمكننا مقارنة شخصين يعيشان في مناطق مختلفة فإننا نتوقع أنهما يحققان مستوياتٍ اقتصاديةً مختلفةً أيضاً. وعليه فإننا لا نتعامل مع الدخل كمتغير متميز كلياً، ولكننا بدلاً من ذلك، ننظر إليه بشكل ما على أنه مرتبط بمكان الإقامة للشخص. ولكي نبين مثل هذه العلاقة في البيانات التي تم جمعها، فإننا سنستخدم الإحصاءات الوصفية التي لا تلخص التوزيع لكل متغير بشكل منفصل فقط، ولكنها أيضاً تصف الطريقة التي تتغير بها قيم أحد المتغيرين والتي ترتبط بتغير القيم في المتغير الآخر.

دعنا نعطي موقفاً مخالفاً عندما يكون لدينا متغيران غير مرتبطين ببعضهما البعض.

بمعنى آخر، أنهما مستقلان. والمتغيران يكونان مستقلين إذا كان نمط التباين في الدرجات لأحدهما ليس مرتبطاً بنمط التباين في الدرجات للمتغير الآخر.

توجد لدينا عدة تقنيات، نطلق عليها إجمالاً الإحصاءات الوصفية الثنائية والتي يمكن أن تستخدم من قِبل الباحث لمعرفة ما إذا كانت هناك علاقة بين متغيرين، انظر الجدول التالي:

جدول (6ـ1) الإحصاءات الوصفية الثنائية

الطريقة	الأمثلة
- طرق الجدولة والتمثيل البياني، هذه الطرق تمدنا ببيانات بطريقة تبين احتمالية العلاقة بين متغيرين.	- الجدول الثنائي لبيانات مقولية "اسمية وترتيبية". رسم انتشاري لبيانات متصلة "المقياس ذو المسافات والنسبي".
- الطرق العددية، هذه الطرق هي عمليات رياضية تستخدم للتكميم لرقم أحادي، قوة العلاقة. وهذه الطرق يطلق عليها قياس العلاقات. وعندما يكون كلا المتغيرين قد تم قياسهما على الأقل على المستوى الترتيبي فهما يشيران أيضاً إلى اتجاه العلاقة.	(مستوى اسمي) Lambda, Cramerśv (قياس ترتيبي بقيم قليلة) Gamma, Kendalś tau-b Somerś d and tau-c (قياسات ترتيبية بقيم كثيرة) Spearmanś rank-orders Correlation coefficient (ذو مسافات ونسبي) Pearsonś correlation coefficient

المصدر: George Argyrous, op. cit. P. 136.

هذه الطرق متممة وتساند بعضها البعض، وهدف كل طريقة من هذه الطرق هو مدنا بالبينة أو الدليل لأي اعتماد قد يوجد بين متغيرين.

نمذجة العلاقة بين متغيرين[1]:

إن ما سوف نتعلمه في هذا الفصل والفصول اللاحقة له هي التقنيات المحددة لاستقصاء العلاقة بين متغيرين، إلا أنه في بادئ الأمر نحتاج إلى أن ننظر حول المستوى التصوري إلى ماذا يعني الارتباط بين متغيرين.

نفترض أننا قمنا بجمع بيانات حول متغيري الدخل ومكان الإقامة. إننا نقوم بإجراء هذا البحث معتقدين أن هناك علاقة بين هذين المتغيرين، آملين أن نكشف خصائص هذه العلاقة في البيانات التي قمنا بجمعها. حيث يمكننا تصوير أو نمذجة العلاقات الممكنة بين هذين المتغيرين – إن وجدت – بطرق متعددة. إن أبسط طريقة يرتبط بها متغيران يمكن أن تسبب العلاقة المباشرة والتي يمكن أن نبينها في الأشكال المحتملة التالية:

1- علاقة مباشرة أحادية الاتجاه "الدخل كمتغير تابع":

إن أنماط هذه العلاقات كطريق واحد يمتد من مكان الإقامة إلى الدخل (انظر شكل 6-1). ويمكن أن تكون لدينا نظرية من خلالها يمكننا القول إن المهنة وفرص الحياة تتباين عبر المدن، وبالتالي يؤثر ذلك على مستوى الدخل للأفراد الذين يعيشون في تلك المدن. في هذه الحالة بالذات يمكننا أن نجادل في أنه يوجد لدينا نمط الاعتماد مع الدخل كمتغير تابع ومكان الإقامة كمتغير مستقل.

مكان الإقامة (متغير مستقل) ← الدخل (متغير تابع)

شكل (6-1) علاقة أحادية الاتجاه "الدخل كمتغير تابع"

2- علاقة مباشرة أحادية الاتجاه "مكان الإقامة كمتغير تابع":

قد تكون هناك مجموعة أخرى من الباحثين الاجتماعيين الذين لا يوافقوننا على النموذج السابق، لأن أولئك الباحثين ينطلقون من بعد نظري آخر يسلم بوجود نمط الاعتماد بين المتغيرين، ولكن هذا النمط يتخذ اتجاهاً آخر. فالناس ذوي الدخول العالية يمكنهم أن يختاروا مكاناً مناسباً للإقامة، وبالتالي قد ينتقلون إلى المدينة ذات البيئة المرغوبة. وعليه فإن مكان الإقامة سيكون متغيراً تابعاً والدخل متغيراً مستقلاً. (انظر شكل 6 – 2).

الإقامة ← الدخل

(متغير تابع) (متغير مستقل)

شكل (6-2) علاقة أحادية الاتجاه "مكان الإقامة كمتغير تابع"

3- علاقة مباشرة ثنائية الاتجاه بين مكان الإقامة والدخل بشكل متعمد تبادليا:

أيضاً قد تكون هناك مجموعة ثالثة من علماء العلم الاجتماعي يمكنهم أن يتفقوا على أن المتغيرين مرتبطان بعضهما ببعض، ولكنهم يعتقدون أن كلا النمطين من السببية (العلّية) يعملان بكيفية تجعل المتغيرين يؤثر بعضهما في الآخر. أي علاقة تبادلية.

في هذا النموذج إنه من غير الملائم تصوير أن واحداً من هذين المتغيرين كمتغير مستقل والآخر كمتغير تابع. ولكن بدلاً من ذلك، يمكن النظر إليهما بأنهما معتمدان على بعضهما البعض تبادلياً.

مكان الإقامة ↔ الدخل

(متغير تابع) (متغير تابع)

شكل (6-3) علاقة ثنائية الاتجاه بين مكان الإقامة والدخل بشكل تبادلي

وعليه فإنه من الأهمية بمكان أن لا يكون هناك تشويش في بيان العلاقة المعتمدة بالعلاقة المتقاربة Closely related ولكن ما يميز المشكلة هو تحديد ما المتغير التابع؟ وما المتغير المستقل. وتحدث هذه العملية فقط في حالة العلاقة أحادية الاتجاه وهي الطريقة المناسبة لتحديد أحد المتغيرين كمتغير مستقل والآخر كمتغير تابع. كذلك من الأهمية بمكان أن نتذكر عند اختيارنا لنماذج تستند على وجهات نظرية محددة حول طبيعة العالم الواقعي وسلوك البشر، فإن هذه النماذج يمكن أو لا يمكن أن تكون صحيحة. فالتحليل الإحصائي لا يبرهن البتة على أي من أنماط السببية التي أوضحناها سابقاً. وكل ما تبينه هذه التحليلات الإحصائية هو أن بعض العلاقات الإحصائية بين المتغيرات المشاهدة قد استندت على بيانات تم جمعها.

إن الطريقة التي تنظم وتفسر بها البيانات تبنى على النتائج التي تتوقف على هذه الافتراضات النظرية المسبقة. وإن نفس البيانات يمكن أن تبين لنا مجموعة من الفروق المعتمدة على افتراضات نظرية لما تمدنا به البيانات. فعلى سبيل المثال، لقد أوضحنا ثلاثة نماذج بسيطة لوصف العلاقة بين متغيرين وتجدر الإشارة إلى أن هناك نماذج أكثر تعقيداً والتي تتضمن العلاقة بين ثلاثة متغيرات أو أكثر. ولاكتشاف العلاقات الأكثر تعقيداً قد يقودنا الأمر للتعامل مع التحليل المتعدد الذي يبحث في العلاقات بين أكثر من متغيرين - سنؤجل الحديث في هذا الأمر إلى الجزء الرابع من هذا الكتاب - وعلى أي حال فإنه من الأهمية بمكان أن نضع في اعتبارنا عند تفسيرنا للنتائج الثنائية الحقيقة القائلة إن أي علاقة مشاهدة بين متغيرين يمكن أن تكون أكثر تعقيداً مقارنة بالنماذج البسيطة المتعلقة بالسبب والنتيجة كما وصفناها سابقاً.

والآن يمكننا أن نبين التقنيات المتعلقة بوصف العلاقات الثنائية وذلك من خلال الجدول الثنائي.

وصف البيانات المقولية باستخدام جداول التقاطع:

دعنا نفترض أننا نقوم بدراسة لاستقصاء العلاقة بين مكان الإقامة ومستوى الدخل، إننا قد نشك في وجود علاقة بين هذين المتغيرين وقد يقودنا هذا الشك إلى أبعد من ذلك، للاعتقاد بأن العلاقة السببية الأحادية تبدأ من مستوى الدخل إلى مكان الإقامة. بمعنى آخر يمكننا أن نجادل في أن مستوى الدخل هو المتغير المستقل، وأن مكان الإقامة هو المتغير التابع. ويتوجب علينا أن نتذكر أنها الطريقة الافتراضية الوحيدة لما نتوقع أن نجده. إن العالم الحقيقي، أو على الأقل البيانات التي تم جمعها من هذا العالم، قد لا تتفق مع هذا التوقع. إن هذين المتغيرين قد لا يرتبطان بقدر ما يكونان مستقلين عن بعضهما البعض. والسؤال هنا هو: كيف يمكن لنا أن ننظم البيانات التي تم جمعها لنبين ما إذا كان النموذج الذي نستخدمه نموذجاً صحيحاً أو ما إذا كان المتغيران في حقيقة الأمر مستقلين عن بعضهما.

إن متغير مكان الإقامة قد تم تعريفه إجرائياً من خلال طرح سؤال على (720) شخصاً بخصوص ما إذا كانوا يعيشون في مناطق ريفية أو حضرية (مقياس اسمي) في حين قد تم تعريف الدخل تعريفاً إجرائياً على مستوى المقياس الترتيبي وذلك من خلال تحديد ما إذا كان الشخص يكسب أكثر أو أقل من متوسط الدخل العام؛ إن هؤلاء الأشخاص الذين يكسبون أقل من المتوسط العام قد تم تصنيفهم إلى ذوي الدخول المنخفضة أما الذين يكسبون أكثر من المتوسط العام، فقد تم تصنيفهم إلى فئة ذوي الدخول العالية. إن هذا البحث الذي نجريه يقدم لنا (720) زوج من الأعداد، وإن كل شخص سوف تخصص له قيمته لتشير إلى المكان الذي يعيشون فيه. وقيمة أخرى تشير إلى المجموعة الاقتصادية التي ينتمي إليها. والسؤال هنا: كيف يمكن لنا أن ننظم هذه الأعداد المتعلقة بـِ (1440) شخصاً بطريقة تظهر لنا أي علاقة موجودة بين الدخل ومكان الإقامة؟. يمكننا في هذه الحالة أن نستخدم الطرق الأحادية التي تعلمناها في الفصول السابقة من هذا الكتاب ولبناء توزيعات تكرارية منفصلة لكل متغير. (جدول 6 – 2 وجدول 6 – 3).

جدول (6-2) التوزيعات التكرارية لمكان الإقامة

مكان الإقامة	التوزيع التكراري
حضري	397
ريفي	323
المجموع	$\sum 720$

جدول (6-3) التوزيع التكراري لمستوى الدخل

مستوى الدخل	التوزيع التكراري
دخل منخفض	372
دخل عالي	348
المجموع	$\sum 720$

إنه من الواضح أن هذه التوزيعات الأحادية المنفصلة في الجدولين السابقين لا تساعدنا البتة كثيراً وذلك لأنه من الصعب أن نلاحظ ما إذا كانت العلاقة موجودة بين متغيرين، والتي هي هدف الدراسة. ولكي نتحصل على أي علاقة محتملة يمكن أن توجد بين متغيرين تم قياسهما ببيانات مقولية فإنه ينبغي علينا استخدام الجدول الثنائي والذي يُعرف أيضاً بجدول التوافق أو جدول التقاطع (Crosstab). فالجدول الثنائي يعرض لنا التوزيعات التكرارية المشتركة لمتغيرين، إن جدول التقاطع لهذه البيانات قد تم جمعهما (نظرياً) يوضحها الجدول (6 – 4).

جدول رقم (6ـ4) العلاقة بين مكان الإقامة ومستوى الدخل

مكان الإقامة	الدخل		المجموع
	منخفض	عال	
حضري	167	230	397
ريفي	205	118	323
المجموع	**372**	**348**	$\sum$720

المصدر: George Argyrous, Statistics for Social & Health Research, op. cit, P. 140.

من خلال جدول التوافق الذي يبين التوزيع المشترك لمتغيرين يمكننا أن نقرأ الدرجات التي أعطيت لكل حالة من حالات المتغير بشكل تلقائي. وبالنظر إلى هذا الجدول على سبيل المثال، يمكننا أن نرى أن هناك 205 حالة تقطن في المناطق الريفية ولديها دخول منخفضة. وبما أن الجداول الثنائية تصف بيانات بطريقة تُظهر هذا التوزيع المشترك، كما أنها تسمح لنا باستقصاء فيما إذا كان هذان المتغيران مرتبطين.

وتجدر الإشارة إلى أن هناك مجموعة من القواعد التي يتوجب على الباحث اتباعها لبناء الجداول الثنائية:

- لابد أن يكون للجدول عنوان ملائم، فجدول التقاطع يجب دائماً أن يكون له عنوان مع بيان قيم كِلاَ المتغيرين.
- الإشارة إلى مصدر البيانات، ويمكن أن يكون المصدر قبل أو بعد الجدول أو هامش مرتبط بالجدول كما هو وارد في الجدول السابق.
- وضع المتغيرات في مكانها المناسب في الصفوف والأعمدة - عادة ما يوضع المتغير المستقل في العمود والمتغير التابع في الصف - إذا كان هناك أي سبب للاعتقاد أن المتغيرين ليسا مرتبطين ببعضهما البعض ولكن أحد المتغيرين تابع على الجانب الآخر

(علاقة اتجاهية واحدة)، فالمتغير المستقل يجب أن يرتب عبر الأعمدة في حين أن المتغير التابع ينظم طبقاً للصفوف. في هذا المثال الذي أمامنا قد خصصنا أن يكون الدخل متغير مستقل (العمود) في حين أن متغير مكان الإقامة متغير تابع (الصف). إننا نقوم بذلك لاعتقادنا، وعلى ضوء النموذج المستخدم الذي مفاده أن دخل الفرد يحدده مدى العوامل التي في النهاية تؤثر على اختيار الناس أين يعيشون. وعلى أي حال، يمكننا فقط أن نـجادل بشكل شرعي أن هناك علاقة عكسية: فالدخل متغير (تابع) ومكان الإقامة متغير (مستقل). إذا ما اعتقدنا في هذا النموذج، فإننا بالتالي يمكننا بناء جدول ثنائي حيث يكون الدخل في الصفوف ومكان الإقامة عبر الأعمدة. وبطريقة أخرى يمكننا أن نـجادل أن الاعتماد يكون متبادلاً وعليه فالقضية لا تعني شيئاً ما إذا كان المتغير في الصفوف أو عبر الأعمدة. ولمناقشة استخدام جداول التقاطع كوسيلة لوصف البيانات، يتوجب علينا أن نكون على دراية ببعض المصطلحات:

- **حجم وأبعاد الجدول**: يعرف حجم الجدول بعدد الفئات لمتغير الصفوف مضروب في عدد الفئات لمتغير العمود. في هذا المثال، لدينا مجموعتان من الفئات هما: مكان الإقامة (حضري-ريفي)، وفئتا الدخل (منخفض- عالٍ) يولدان لنا جدول 2×2. وإذا تم قياس الدخل على مقياس يحتوي على أربع نقاط مثل: (دخل منخفض جداً، منخفض، عالٍ، عالٍ جداً) فإننا بالتالي سنتحصل على جدول تكون أبعاده 2×4.

- **خلايا الجدول**: كل مربع في الجدول سيحتوي على عدد الحالات التي لديها مجموعة متآلفة من القيم لمتغيرين، وفي هذه الحالة نطلق عليها جدول الخلية.

- **هوامش الجدول:** إن مدونات مجموع الأعمدة يطلق عليها هوامش العمود وبنفس الكيفية فإن مدونات مجموع الصفوف يطلق عليها هوامش الصف.

جداول التقاطع ذات التوزيعات النسبية [2] :

بإمكان الباحث أن يطوِّر قدراته ليرسم أي علاقة محتملة موجودة في البيانات وذلك بحساب التوزيعات النسبية، بدلاً من الأعداد المطلقة من الحالات في كل فئة مشتركة. وفي كل خلية من خلايا الجدول نبين المعلومات في إطار النسب إما لمجاميع الأعمدة أو لمجاميع الصفوف. إن التوزيعات النسبية تقوم أساساً على مجموع الأعمدة كما هو موضح في الجدول التالي:

جدول (6ـ5) العلاقة بين مكان الإقامة والدخل (N = 720)

مكان الإقامة	الدخل		المجموع
	منخفض	عال	
حضري	$\frac{167}{372} \times 100 = 45\%$	$\frac{230}{348} \times 100 = 66\%$	$\frac{397}{720} \times 100 = 55\%$
ريفي	$\frac{205}{372} \times 100 = 55\%$	$\frac{118}{348} \times 100 = 34\%$	$\frac{323}{720} \times 100 = 45\%$
المجموع	$\frac{372}{372} \times 100 = 100\%$	$\frac{348}{348} \times 100 = 100\%$	$\frac{720}{720} \times 100 = 100\%$

إن قيمة 55% تمثل عدد أولئك الذين يعيشون في المناطق الريفية ولديهم مستويات اقتصادية منخفضة وهي نسبة تمثل العدد الكلي لأولئك الذين يكسبون الدخول المنخفضة:

$$\frac{205}{372} \times 100 = 55\%$$

وكما يمكننا أيضاً، وبشكل آخر أن نبين التوزيعات النسبية باستخدامنا لمجموع الصفوف كما هو مبين في الجدول رقم (6) الذي يحتوي على الحسابات المتعلقة بالقاطنين في المناطق الحضرية. ويمكننا بوجه السرعة أن نرى من هذه العملية أن 42 % من الناس كنسبة فقط من العدد الكلي لأولئك الناس الذين يعيشون في المناطق الحضرية ولديهم مستوى اقتصادي منخفض. وفي بعض الأحيان فإن الفرق في التوزيعات التكرارية للجداول يمكننا أن نجمعها لتعطينا البيانات الخام والنسب المئوية المناسبة وذلك بإضافة أعمدة أو صفوف أخرى.

جدول (6ـ6) العلاقة بين مكان الإقامة والدخل (N = 720)

مكان الإقامة	الدخل		المجموع
	منخفض	عال	
حضري	$\frac{167}{397} \times 100 = 42\%$	$\frac{230}{397} \times 100 = 58\%$	% 100
ريفي	% 63	% 37	% 100
المجموع	**% 52**	**% 48**	**% 100**

إن أنسب بناء يعتمد على محتوى البيانات التي تهم المشاهدين وعلى أية حال، وبالرغم من المجاميع التي من خلالها يتم حساب التوزيعات النسبية، فإنه من الأهمية بمكان، أن نضيف حجم العينة الحقيقي حتى يمكننا إعادة تحويل النسب إلى الأرقام المطلقة إذا ما طُلب منا ذلك.

إجراءات توليد جداول التقاطع Cross tables باستخدام SPSS:

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Analyze ⟸ Descriptive Statatics ⟸ Crosstabs

2- انقر على المتغير في القائمة التي تشكل صفوف الجدول والتي تكون في هذه الحالة Place of residence.

3- انقر على ◄ التي تشير إلى القائمة المحددة المُعَنْوَنَة Row(s) نقوم بلصق Place of residence في القائمة المحددة للصفوف Row(s).

4- انقر على المتغير في القائمة التي تشكل أعمدة الجدول والتي تكون في هذه الحالة Income Level.

5- انقر على ◄ التي تشير إلى القائمة المحددة المُعَنْوَنَة Column(s) نقوم بلصق Income Level في القائمة المحددة للأعمدة Column(s).

6- انقر فوق OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

Place of residence Income Level Crosstabulation

Count

		Income Level		TOTAL
		Low	High	
Place of residence	urban	167	230	397
	Rural	205	118	323
TOTAL		**372**	**348**	**720**

شكل (6-4) مخرجات Crosstabs باستخدام Spss

يمكن للباحث من خلال استخدام برنامج Spss أن يتوسع في توليد جدولٍ يحتوي على التوزيعات النسبية relative والمطلقة absolute.

كما يظهر في الشكل التالي:

Place of residence Income Level Crosstabulation

		Income Level		TOTAL
		Low	High	
Place of residence	urban Count	167	230	397
	% Within Place of residence	42.1 %	57.9 %	100.0 %
	% Within Income Level	44.9 %	86.1 %	55.1 %
	Rural Count	205	118	323
	% Within Place of residence	63.5 %	36.5	100.0 %
	% Within Income Level	55.1 %	33.9 %	44.9
TOTAL	Count	372	348	720
	% Within Place of residence	51.7 %	48.3 %	100.0 %
	% Within Income Level	100.0 %	100.0 %	100.0 %

المصدر: George Argyrous, op.cit., PP. 143 - 144

شكل (6-5) مخرجات SPSS لجداول التقاطع بالتوزيعات النسبية

تجدر الإشارة إلى أنه من خلال جدول واحد يمكن تقديم كل المعلومات التي تم توليدها بشكل منفصل في الجداول 4، 5، 6.

تفسير جداول التوافق: نمط وقوة العلاقة:

يمكننا في هذه الجزئية أن نقدم أداة مهمة لوصف جداول التوافق في البحوث الاجتماعية وأهمية هذه الأداة تكمن في أننا في مجال البحث الاجتماعي نقوم بجمع بيانات كثيرة وهي بيانات منفصلة وبيانات مقولية مقاسة على مستوى المقياس الاسمي أو الترتيبي. وعندما تُحول مجموعة من البيانات الخام إلى جدول التقاطع فإن المهمة تكمن في كيفية تفسيرها. ولتقيِّيم ما إذا كانت هذه البيانات تُظهر لنا أن هناك علاقة موجودة بين المتغيرين. وعند تفسير أي علاقة واضحة في جدول التوافق فإننا في واقع الأمر ننظر إلى خاصيتين أساسيتين:

1- نمط العلاقة.

2- قوة العلاقة.

ولكي يتبين شكل العلاقة يبقى من الأهمية بمكان أن نسلط الضوء على خلية النموذج Model Cell لكل عمود، انظر جدول رقم (6 – 7).

جدول رقم (6ـ7) العلاقة بين مكان الإقامة وفقا لمستوى الدخل

مكان الإقامة	الدخل		المجموع
	منخفض	عال	
حضري	167 (%45)	230 (%66)	397
ريفي	205 (%55)	118 (%34)	323
المجموع	**372**	**348**	**720**

وعند إلقاء الضوء على خلية النموذج لكل خلية في العمود يمكننا القول إن هناك علاقة. وبالنظر إلى التوزيعات النسبية فإنه من الواضح أن 55 % من الذين يكسبون

دخولاً منخفضة يعيشون في مناطق ريفية، في حين نجد أن نسبة 66 % من الذين يكسبون دخولاً عالية يعيشون في مناطق حضرية. ومن هنا يمكننا أن نقترح تفسيراً مفاده أن هناك علاقة بين الدخل ومكان الإقامة، ويشير نمط هذه العلاقة إلى أن أولئك الذين يكسبون دخلاً منخفضاً يتركزون في الريف، في حين إن أولئك الذين يكسبون دخولاً عالية يتركزون في المدن.

كذلك يمكننا أن نقيّم قوة هذه العلاقة وذلك بالنظر إلى نسبة الحالات في كل عمود تحصلنا عليها من خلال خلية النموذج في كل عمود. كذلك يمكننا أن نرى أنه بينما فئة النموذج للذين يكسبون دخولاً منخفضة يعيشون في الريف، كذلك توجد نسبة عالية من ذوي الدخول المنخفضة يعيشون في مناطق أخرى.

كذلك الحال، بينما نجد أن الغالبية من الذين يكسبون دخولاً عالية يعيشون في مناطق حضرية، فإن هناك نسبة كبيرة منهم لا يعيشون في هذه المناطق الحضرية، وتُبَيِّنُ لنا هذه النتيجة أن العلاقة بين الدخل ومكان الإقامة علاقة ليست قوية.

تفسير الجداول المتقاطعة عندما يكون لدينا متغيران مقاسان على المستوى الترتيبي:

في الجزء الأول من هذا الفصل تم التركيز على مناقشة بناء جداول التوافق والكيفية التي يمكن بها تفسير العلاقة في مثل هذه الجداول. ولقد تناولنا مثالاً يكون فيه أحد المتغيرين مقاساً على المستوى الاسمي (مكان الإقامة) ومتغير آخر تم قياسه على المستوى الترتيبي (الدخل). إن القواعد والإجراءات التي تعلمناها من هذا المثال السابق يمكن تطبيقها في بناء جدول التوافق لمتغيرات يمكن قياسها على أي مستويات مؤتلفة. عندما يقاس كلا المتغيرين على المستوى الاسمي، ومع ذلك، فإن التفسيرات في نمط العلاقة الموجود في جدول التوافق يمكن أن يتخذ خطوة أبعد ليدمج في مناقشة علاقة الاتجاه والاتساق.

1- اتجاه العلاقة Direction of Relationship:

تحدثنا في السابق عن العلاقة بين الدخل ومكان الإقامة وتوصلنا إلى نتيجة مفادها أن الدخل يرتبط بمكان الإقامة كأن يميل ذوو الدخول المحدودة أكثر إلى السكن في الأماكن

الريفية، بينما يميل ذوو الدخول العالية أكثر إلى العيش في المناطق الحضرية وجاءت هذه النتيجة لأننا تعاملنا مع متغير (مكان الإقامة) الذي تم قياسه على المستوى الاسمي وبالتالي لا نستطيع أن نتحدث عن الزيادة أو النقصان في الدخل من حيث كونه مرتبطاً بالزيادة أو النقصان في مكان الإقامة. وحسبنا أن نشير إلى أنه من غير المنطقي الحديث حول مكان الإقامة من حيث كونه أكبر أو أصغر فمكان الإقامة يتباين عبر الحالات ولكنه لا يزيد أو ينقص كمياً.

وعندما يقاس كلا المتغيرين على الأقل على المستوى الترتيبي فإنه بإمكاننا أن نتحدث عن اتجاه العلاقة من حيث كونها سالبة أو موجبة. فالعلاقة الموجبة تعني أن الزيادة في درجات وحدة المقياس (x) يقابلها زيادة في وحدة المقياس بالنسبة لمتغير (y) في حين إن العلاقة السالبة تعني أن الزيادة في درجات وحدات المقياس بالنسبة لمتغير (x) يقابلها نقص في وحدة المقياس بالنسبة لمتغير (y). فعلى سبيل المثال، قد نرغب في معرفة العلاقة بين الدخل ومشاهدة الإذاعة المرئية لشخص ما، بدلاً من العلاقة بين الدخل ومكان الإقامة، إن كمية ما يشاهده هذا الفرد يمكن قياسها من خلال طرح السؤال التالي: ما إذا كان الشخص يشاهد الإذاعة المرئية، أو لا يشاهدها البتة، يشاهدها بعض الليالي، أو معظم الليالي.

إن مثل هذا السؤال تم قياسه على المستوى الترتيبي أي أن كلا المتغيرين الآن تم قياسهما على المستوى الترتيبي، فإذا ما استطعنا أن نجد علاقة موجودة فعلاً في هذه الحالة يمكننا الحديث عن اتجاه العلاقة. نفترض أننا قمنا بجمع معلومات حول 300 شخص وتم قياس دخل هؤلاء الناس، (هذه المرة تم قياس الدخل على مقياس يتكون من ثلاث نقاط: دخل منخفض - دخل متوسط - دخل عال). وكمية مشاهدة الإذاعة المرئية، بإمكاننا بشكل أولي أن نقلص 600 حالة وهي تمثل البيانات الخام للبحث إلى جدول ثنائي. إننا نشك إذا ما وجد نمط الاعتماد بين هذين المتغيرين الذي سوف يبدأ من الدخل إلى مشاهدة الإذاعة المرئية. وعليه يمكننا أن نرتب الجدول على أن يكون الدخل (متغير مستقل) عبر الأعمدة ومشاهدة الإذاعة المرئية (كمتغير تابع) على الصفوف.

إن كل القواعد التي تعلمناها في بداية هذا الفصل والمتعلقة ببناء جداول التوافق لازالت قابلة للتطبيق. إلا أن هناك قاعدة إضافية مهمة عندما يقاس المتغيران على المستوى الترتيبي. وعندما يحتوي جدول التوافق على متغيرين مقاسين على المستوى الترتيبي، يستوجب ذلك علينا ترتيب الجداول بطريقة تزيد فيها قيم المتغير المستقل عبر الصفحة من اليمين إلى اليسار في حين أن قيم المتغير التابع تنقص أسفل الصفحة.

جدول رقم (6ـ8) توزيع مشاهدة الإذاعة المرئية حسب الدخل

مشاهدة الإذاعة المرئية	الدخل			المجموع
	منخفض	متوسط	عال	
أبداً	75 ٪71.5	15 ٪15	10 ٪10.5	100
بعض الوقت	20 ٪19.0	70 ٪70	10 ٪10.5	100
معظم الوقت	10 ٪9.5	15 ٪15	75 ٪79.0	100
المجموع	**105** **٪100.0**	**100** **٪100.0**	**95** **٪100.0**	**300**

ولكي نستطيع تفسير الجدول كما تم في المثال السابق يستوجب ذلك علينا أن نلقي الضوء على خلية النموذج لكل عمود وبالتالي يمكننا أن نرى بشكل سريع أن هناك علاقة بين هذين المتغيرين، أي أن الزيادة في الدخل يقابلها زيادة في كمية مشاهدة الإذاعة المرئية. ومن هنا يمكننا القول بأن هناك علاقة موجبة. وإذا كانت خلايا النموذج كلها تتراصف على طول القطر الآخر من الأعلى إلى أبداً إلى منخفض إلى معظم الليالي فإن هذا الجدول سيصف علاقة سالبة.

2- علاقة الاتساق Consistency of the Relation:

بالإضافة إلى اتجاه العلاقة عندما تعاملنا مع المتغيرين اللذين تم قياسهما على المستوى الترتيبي، فإنه بالتالي يمكننا أن نبحث فيما إذا كانت العلاقة علاقة اتساق. لاحظ أن كل خلايا النموذج في جدول (8) قد رتبت على طول القطر الموجب، وبالتالي يوجد تعاقب في العلاقة عبر كل مدى القيم، إن مثل نمط الاعتماد هذا يطلق عليه "علاقة الاتساق". إلا أنه على الطرف الآخر يمكننا مشاهدة النتيجة التي يحتويها الجدول رقم (6 – 9).

ومن خلال بيانات هذا الجدول يمكننا إعادة القول بأن هناك علاقة بين المتغيرين ويمكن وصفها بأنها علاقة ليست متسقة.

جدول رقم (6-9)

العلاقة بين مشاهدة الإذاعة المرئية حسب الدخل (علاقة غير متسقة)

مشاهدة الإذاعة المرئية	الدخل			المجموع
	منخفض	متوسط	عال	
أبداً	75 %71.5	10 %10.0	15 %15	100
بعض الوقت	20 %19.0	10 %10.0	70 %70	100
معظم الوقت	10 %9.5	75 %71.5	15 %15	100
المجموع	**105** **%100.0**	**95** **%100.0**	**100** **%100.0**	**300**

من خلال هذا الجدول يمكننا أن نلاحظ في نهاية الفقرة المتعلقة بقياس الدخل، أنه كلما زاد الدخل زادت معه كمية مشاهدة الإذاعة المرئية. لكننا نلاحظ أن العلاقة بين المتغيرين هي علاقة عكسية، كلما تحركنا أبعد على فقرات مقياس الدخل.

الخلاصة:

لقد حاولنا في هذا الفصل أن نبين بشكل واسع بناء وتفسير الجداول الثنائية. ومن خلال هذا الفصل تبين لنا أهمية هذه الجداول في وصف البيانات المقولية بطريقة تكشف لنا ما إذا كانت هناك علاقة بين متغيرين يخضعان للاستقصاء؛ كما ناقشنا في هذا الفصل القواعد المحددة والإجراءات لتحويل البيانات الخام المجمعة إلى جداول تقاطعية مندمجة. كما بيّنّا الأسس التي على ضوئها يمكن تفسير أي علاقة يبينها جدول التوافق، وذلك كما اتضح لنا في كل الجداول التي حاولنا من خلالها تقييّم نمط وقوة العلاقة؛ أيضاً حاولنا من خلال هذا الفصل قياس متغيرين مقاسين على المستوى الترتيبي لإعطائنا أيضاً أبعاداً إضافية للعلاقة التي يمكن أن نكشفها باستخدام جداول التوافق، أعني اتجاه العلاقة وما إذا كانت العلاقة متسقة أم لا.

أسئلة للمراجعة:

1- من الجداول التالية: احسب النسب المتعلقة بالعمود:

أ:

التابع	المستقل		المجموع
	1	2	
1	30	60	90
2	45	50	95
المجموع	**75**	**110**	**185**

ب:

التابع	المستقل			المجموع
	1	2	3	
1	56	40	10	106
2	15	30	50	95
المجموع	**71**	**70**	**60**	**201**

2- من كل جدول من الجداول التالية: احسب النسبة المئوية للصف:

أ:

التابع	المستقل		المجموع
	1	2	
1	30	60	90
2	45	50	95
المجموع	75	110	185

ب:

التابع	المستقل			المجموع
	1	2	3	
1	56	40	10	106
2	15	30	50	95
المجموع	**71**	**70**	**60**	**201**

3- عينة افتراضية لأطفال من أستراليا، كندا، سنغافورة وبريطانيا تم مقارنتهم فيما يتعلق بكمية مشاهدة الإذاعة المرئية:

كمية مشاهدة الإذاعة المرئية	البلد				
	كندا	أستراليا	بريطانيا	سنغافورة	المجموع
منخفضة	23	25	28	28	104
متوسطة	32	34	39	33	138
عالية	28	30	40	35	133
المجموع	83	89	107	96	375

المطلوب: هل يمكنك القول إن كمية مشاهدة الإذاعة المرئية (التلفزيون) مرتبط ببلد الإقامة؟

4- عينة مؤلفة من 162 رجلاً بين 40 و 65 سنة تم اختيارهم لمعرفة الوضع الصحي لكل رجل. وتم طرح السؤال التالي على كل منهم لمعرفة ما إذا كان يدخن بشكل روتيني. وجاءت نتيجة هذه العينة بالشكل التالي:

الوضع الصحي	عادة التدخين		المجموع
	يدخن	لا يدخن	
سيء	34	13	47
متوسط	19	22	41
جيد	9	35	44
جيد جداً	3	27	30
المجموع	**65**	**97**	**162**

المطلوب:

- بيّن المتغير المستقل والمتغير التابع ومستوى قياس كل منهما؟
- هل يمكنك تصور أي علاقة محتملة بالنظر إلى أحد المتغيرين كمتغير تابع والآخر كمتغير مستقل، مع تبرير إجابتك؟
- من هذا الجدول احسب نسب العمود.

الهوامش والمصادر:

أولا: الهوامش:

1- George Argyrous, Statistics for Social and Health Research With a Guide to SPSS, Sage Publications, London ,2001, pp. 137 - 141.

2- 2. Ibid, PP. 141 - 149.

ثانيا: المصدران الرئيسيان لهذا الفصل :

1- George Argyrous, Statistics for Social and Health Research With a Guide to SPSS, Sage Publications, London, 2001 .

2- Joseph F. Healey, The Essentials of Statistics: A Tool for Social Research, Wadsworth Cengage Learning, USA, 2010.

الفصل السابع

الوصف العددي للبيانات الاسمية :
مقاييس التطابق الثنائية

مقدمة

في الفصول السابقة أوضحنا الكيفية التي من خلالها تُبنى جداول التقاطع التي تُعْرَف بأنها: الطريقة التي تُنَظَّمُ بها البيانات المقولية بطريقة تكشف لنا ما إذا كانت هناك علاقة بين المتغيرات. ولتوضيح ذلك نسوق المثال التالي:

جدول (7ـ1) مكان الإقامة وفقا لمستوى الدخل

مكان الإقامة	الدخل		المجموع
	منخفض	عال	
حضري	167 (%45)	230 (%66)	397 (%55)
ريفي	205 (%55)	118 (%34)	323 (%45)
المجموع	**372 (%100)**	**348 (%100)**	**720 (%100)**

المصدر: George Argyrous, op.cit, p. 152

عند تحليل جداول التقاطع ننظر ما إذا كانت هناك علاقة موجودة ونهدف من وراء جداول التقاطع إلى طرح سؤالين مرتبطين:

1- ما هو نمط هذه العلاقة؟

2- وكم قوة هذه العلاقة؟

يمكننا أن نرى في هذا الجدول نمط العلاقة التي تشير إلى أولئك الذين يكسبون دخلاً منخفضاً نسبياً يعيشون في مناطق ريفية في حين يعيش أولئك الذين يكسبون دخلاً عالياً نسبياً في مناطق حضرية.

بمعنى آخر، الزيادة في الدخل تتجه إلى أن ترتبط بتغير مكان الإقامة من الريف إلى المناطق الحضرية. كذلك يمكننا أيضاً وصف هذه العلاقة بمصطلحات لفظية: كالعلاقة القوية. فالتباين في الدخل الأدنى إلى الأعلى يرتبط بكمية محددة من التباين في مكان الإقامة، من الريف إلى الحضر، ولكن هذا التباين ليس بكمية كبيرة من التغير.

وتجدر الإشارة إلى أنه قد نكون أكثر موضوعية إذا ما كانت لدينا طريقة لقياس قوة العلاقة في جداول التقاطع بدلاً من أن نعتمد على العين المجردة للحكم فيما إذا كان هناك تباين من شخص لآخر، لذلك فإنه من الأفضل أن تكون لدينا طريقة لقياس قوة العلاقة، لكي تمدنا بنفس الإجابة، بغض النظر عما يدليه الشخص من حكم. وهذا بالتحديد الوظيفة الأساسية لمقاييس التطابق. مثال، قد يختلف اثنان في حكمهما اليوم فيما إذا كان الجو حاراً نسبياً. فشخص يرى أن الجو حار نسبياً، في حين شخص آخر قد يحكم بأن الجو حار جداً، بينما شخص ثالث لديه شعور بأن درجة الحرارة تميل إلى البرودة. كلنا أشخاصاً نمر بنفس الخبرة المتعلقة بدرجة حرارة هذا اليوم، إلا أن انطباعاتنا الذاتية لهذه الخبرة تكون مختلفة. إذا ما نظرنا على أية حال، فكلنا نشير إلى مقياس الحرارة وننظر إلى درجة الحرارة التي وصلت إلى 20 درجة. هذا الأمر قد نوافق عليه جميعاً وإن مقياس درجة الحرارة يبين لنا نفس الدرجة بِغضّ النظر عمن ينظر إليه. فمقياس درجة الحرارة مقياس موضوعي باعتباره مقياس مكمم يستند على معيار عام.

وبنفس الطريقة قد نجد مجموعة مختلفة من الناس يمكن أن تنظر إلى جدول التقاطع

ولفظياً تُقيّم قوة العلاقة بطرق مختلفة، فمقاييس التطابق يمكن أن تمدنا بمؤشر حاسم لقوة العلاقة التي سوف تعطي نفس الإجابة لكل شخص[1].

1- مقاييس التطابق كإحصاءات وصفية:

تُعرف مقاييس التطابق: بأنها إحصائيات وصفية تقيس العلاقة بين المتغيرات. كما تشير إلى مفاهيم كمية وإلى أي مدى أن التغير في قيمة أحد المتغيرات يرتبط بالتغير في قيمة المتغير الآخر. بمعنى آخر يشير التطابق إلى "الاعتماد" عندما تزيد سنوات العمر هل يزيد أيضاً الطول أو يقل؟ هل التغير في المعتقدات الدينية يتطابق مع التغير في الاتجاه والمواقف حول عقوبة الإعدام؟ وكما نعرف أن العلاقات قد نحصل عليها من خلال الرسوم البيانية والجداول التي تبين لنا بطريقة ما العلاقة التي يمكن أن توجد بين متغيرين. كذلك باستطاعتنا، بالإضافة إلى هذه الطرق الوصفية البسيطة أن نحسب مقاييس التطابق للتكميم الفعلي للانطباعات التي تحصلنا عليها من خلال هذه الأدوات.

إن الأمر المهم الذي ينبغي أن نتذكره حول مقاييس التطابق أن هذه العلاقة تساعدنا في وصف البيانات، بدلاً من الركون إلى الانطباعات المرئية لجداول التقاطع أو الرسوم البيانية، هذه الأدوات يمكنها في أحسن الظروف أن تمدنا برقم مفرد لقوة ونمط التطابق.

إن المشكلة مع هذه المقاييس هي تحديد الظروف الملائمة التي يمكن أن تقدم بها هذه المعلومات. فإذا كانت الظروف الملائمة لا تنطبق فحينئذٍ يمكن أن تكون هذه المقاييس العددية مقاييس مضللة، وهنا ينبغي أن تعتمد على الرسوم البيانية والجداول المرتبطة بالوصف اللفظي الملائم لوصف علاقة.

ولوضع مفهوم التطابق في الواقع العملي تكون مشكلة غامضة فالعمل مع مقاييس التطابق يمكن أن تكون تجربة غير مجدية وذلك لوجود عدد كبير من المقاييس للاختيار منها، وكلٌ لها فرادتها وقصورها، وفي الغالب فإن هذه المقاييس قد لا تقود إلى نفس النتيجة. فعلى سبيل المثال، أن الكثير من مقاييس العلاقة تكون حساسة فيما يتعلق بالقرار الذي يتم به تصميم المتغير كمتغير مستقل وما هو المتغير التابع. مثل هذه المقاييس

هي مقاييس غير متماثلة Asymmetric. فالمقاييس غير المتماثلة تكون مفيدة عندما نعتقد أن العلاقة لأحد المتغيرات الذي يعتمد على متغير آخر. ولكن على الجانب الآخر، فإنه إذا ما كان لدينا شك حول العلاقة على أنها علاقة متبادلة، فإننا في هذه الحالة نستخدم مقاييس التماثل Symmetric التي تأخذ نفس القيمة بغض النظر عن المتغير الذي تم تحديده بأنه متغير مستقل أو الذي تم تحديده بأنه متغير تابع. الجدول التالي يعطينا بعض التوجهات للاختيار بين المقاييس الأكثر شيوعاً، التي سنتناولها بالتفصيل في الفصول اللاحقة. إن نقطة البداية للقياس هي تحديد المستوى الذي على ضوئه يتم قياس المتغير.

جدول (7-2): مقاييس التطابق Meaoures of association

مستوى القياس	مقاييس التطابق
الاسمي	Lambda لامبيدا Goodman and Krushal tau جودمان - كروشكل تاو Cramerś V كراميرزV
الترتيبي	Somerś d سومرز d Gamma جاما Kendallś tau-b كندلز تاو b Kendallś tau-c كندلز تاو c Spermanś rank order ارتباط سبيرمان Correlation Coefficient
ذو المسافات/ النسبي	Pearsonś(r) ارتباط بيرسون (r) Product-moment Correlation Coefficient

وتجدر الإشارة إلى أنه عندما يقاس متغيران على مستويات مختلفة، فإنه كقاعدة عامة، فإن اختيار القياس يعتمد على أدنى اثنين من مستويات القياس، على سبيل المثال، إذا ما أردنا أن نستقصي فيما إذا كانت هناك علاقة بين النوع "اسمي" والرضا الوظيفي "ترتيبي" فإن أدنى مستوى القياس لهذين المتغيرين هو المقياس الاسمي، وبالتالي يجب

علينا أن نتقيد بمدى المقاييس التي يمكننا حسابها من تلك المقاييس التي أوردناها في الجدول السابق والتي تلي المقياس الاسمي. ولبناء مقياس التطابق فإنه من المفضل مراعاة الخصائص التالية [2]:

1- من الضرورة بمكان أن تأخذ مقاييس التطابق قيمة 1 (أو –1 حيثما كان ضرورياً) في مواقف يكون فيه التطابق كاملاً. ولسوء الحظ ليس الأمر كذلك، وأن ذلك يسبب الكثير من الإحباط المرتبط باستخدام مقاييس التطابق، فبعض مقاييس التطابق يمكن أن تأخذ قيماً أكبر من 1، بينما مقاييس أخرى مثل (جاما) يمكن أن تأخذ قيمة 1 عندما لا يوجد تطابق كامل.

2- إنه من الضرورة أيضاً أن تأخذ مقاييس التطابق قيمة 0 في مواقف لا يوجد فيها تطابق. ولسوء الحظ ليست كل المقاييس تنطبق عليها هذه الخاصية المثالية. إن بعض المقاييس يمكن أن تأخذ قيمة (0) حتى عندما يوجد تطابقُ بيّن وجلّيٌ للعين المجردة.

3- عندما يقاس كلا المتغيرين على الأقل على المستوى الترتيبي فإن علامة + أو – يجب أن تشير إلى اتجاه التطابق: فيما إذا كانت الزيادة في قيمة أحد المتغيرَيْن متطابقة مع الزيادة "تطابق موجب" أو النقصان "تطابق سالب" في قيمة المتغير الآخر.

فيما يتبقى من هذا الفصل سوف نتناول بالنقاش مقاييس التطابق لمتغيرين عندما يكون أحدهما أو كلاهما قد تم قياسه على المستوى الاسمي. وقبل فعل ذلك، إنه من الأهمية بمكان أن نضع في اعتبارنا أن كل هذه المقاييس تشير إلى التطابق، ولكنها ليست بالضرورة أن تبين ما إذا كان أحد هذه المتغيرات يسبب التغير في متغير آخر. فقد يشك الباحث نظرياً أن أحد المتغيرين يسبب التغير في متغير آخر، إلا أن الإحصاء الذي سوف نتعلمه لا يبرهن على مبدأ السببية، ولكن الأمر يتعلق فقط بتقديم شواهد تعزز الافتراضات النظرية. على سبيل المثال يمكننا أن نلاحظ العلاقة بين عدد طيور اللاقلاق في منطقة معينة ومعدل التكاثر في تلك المنطقة، ويمكننا أن نحسب القياس الذي يكمم

هذه العلاقة الإحصائية. إلا أنه ليس باستطاعتنا أن ننتقل من العملية الإحصائية إلى النتيجة التي مفادها أن طيور اللاقلاق قد تسبب معدل التكاثر *.

2- مقاييس التطابق للمتغيرات الاسمية:

وكما بينا على التو مقياس التطابق فإنه باستطاعتنا أن نفكر في مؤشر عددي يشير إلى قوة العلاقة. وهذه المقاييس تمتد ما بين طرفَيْن أحدهما يشير إلى حالة التطابق التام. والآخر إلى عدم وجود تطابق. وباستخدامنا للمثال المتعلق بالدخل ومكان الإقامة. يمكننا بيان ما إذا كان لدينا متغيران متطابقان بشكل تام.

جدول (7-3): مكان الإقامة وفقاً لمستوى الدخل

مكان الإقامة	الدخل		المجموع
	منخفض	عال	
حضري	0	397	397
ريفي	323	0	323
المجموع	**323**	**397**	**720**

يتبين لنا من هذا الجدول أنه إذا كان هناك أشخاص لديهم دخل منخفض فإن ذلك يتيح لنا أن نتوقع بشكل قاطع أنهم يعيشون في مناطق ريفية. فالدخل لهذه المجموعة من الحالات هو متنبئ تام لمكان الإقامة: فمعرفة مستوى الدخل لشخص ما يسمح لنا بالتكهن بشكل تام أين يعيش هذا الشخص. ولكن بطريقة أخرى، فإن التغير من دخل منخفض إلى دخل عال دائماً سوف يتطابق مع تغير في مكان الإقامة من الريف إلى الحضر. ومن خلال التطابق التام يمكننا القول إن كل تباينٍ في المتغير التابع (مكان

* طيور اللاقلاق: هي مجموعة متنوعة من الطيور التي تغوص في الماء بحثاً عن الطعام، من عائلة Ciconiidae وتكثر خصوصاً في نصف الكرة الشرقي وتمتاز هذه الطيور بأرجل طويلة ومناقير مستقيمة.

الإقامة) تم تفسيره بواسطة التباين في المتغير المستقل (الدخل): فالفرق بين هاتين الحالتين فيما يتعلق بمكان الإقامة يمكن تفسيره بمجرد الإشارة إلى الفرق في دخول هؤلاء الأشخاص. وعلى النقيض من ذلك فإن الجدول التالي يتعلق بحالة لا يوجد بها تطابق.

جدول (7ـ4): لا يوجد تطابق

مكان الإقامة	الدخل		المجموع
	منخفض	عال	
حضري	178 (%55)	219 (%55)	397
ريفي	145 (%45)	178 (%45)	323
المجموع	**323**	**397**	**720**

إن التباين في الدخل في هذا المثال، يشير إلى أنه لا توجد علاقة للتباين في مكان الإقامة. فقد جاءت النسبة متساوية لدى الدخول المنخفضة للذين يعيشون في المدينة كما هو الحال لأولئك الذين لديهم دخول عالية. وكل واحد من هذه الفئات للمتغير المستقل هي بالضبط مساوية لنفس نمط الإجابات الموجودة في المتغير التابع. إن هاتين الحالتين اللتين تشيران إلى عدم التطابق والتطابق التام من المتغيرين اللذَيْن يقعان بشكل معاكس في نهاية المقياس. فإن حالة لا تطابق أعطيت قيمة 0 وإن علاقة التطابق التام قد أعطيت قيمة 1. انظر الشكل التالي:



شكل (7ـ1) ميزان مقاييس التطابق الاسمية

إننا في واقع الأمر لا يمكننا جمع بيانات تكون مطابقة لأحد هذين الطرفين المتطرفين. إلا أنهما ببساطة يقومان بمهمة نقاط مرجعية. فالبيانات عادة تقع في مكان ما بين المثال الذي سنقوم بتوضيحه.

جدول (7ـ5) مكان الإقامة وفقاً لمستوى الدخل: تكرارات مشاهدة

مكان الإقامة	الدخل		المجموع
	منخفض	عال	
حضري	167	230	397
ريفي	205	118	323
المجموع	**372**	**348**	**720**

وبالنظر إلى جدول التطابق يتبين لنا وجود بعض العلاقة بين هذين المتغيرين ولكنه من الوضوح أيضاً أن هذه الحالة ليست حالة تطابق تام. فإذا أردنا أن نعطي العلاقة القوية في هذا الجدول عدد يكون بين 0 و 1، فالصفر (0) يشير إلى عدم وجود تطابق في حين يشير رقم (1) إلى تطابق تام.

إن حساب قيمة لامبيدا يعطينا الموقع للعدد الدقيق على طول المتصل في الشكل (1) بحيث نعرف أن النتيجة الفعلية تقع على طول المتصل، ونقوم بذلك من خلال قياس المسافة الإحصائية بين الجدول الذي يحتوي على البيانات الفعلية التي نشاهدها، وأن كل واحد من الاحتمالين للمواقف المتطرفة لعدم وجود تطابق ووجود تطابق تام.

لامبيدا هي واحد من مجموعة المقاييس التي يطلق عليها نسبة التخفيض في الخطأ Proportional reduction in Error (PRE).

منطقة نسبة التخفيض في الخطأ (PRE):

إن منطق هذه المقاييس يتطلب منا أن نقوم بنوعين مختلفين من التنبؤ حول درجات الحالات المدروسة. إن أول هذه التنبؤات ينبغي علينا أن نتجاهل المعلومات المتعلقة بالمتغير المستقل، في هذه الحالة سنقوم بعدة أخطاء بالتنبؤ بالدرجة على المتغير التابع. أما التنبؤ الثاني، فإننا نأخذ في الاعتبار درجة الحالة على المتغير المستقل لمساعدتنا في التنبؤ بالدرجة على المتغير التابع. فإذا كان هناك تطابق بين المتغيرين، فإننا بالتالي نولّد أخطاء قليلة عندما نأخذ في الاعتبار المعلومات المتعلقة بالمتغير المستقل.

إن نسبة التخفيض في الخطأ (PRE) لمقاييس التطابق تعبر عن نسبة التخفيض في الخطأ بين اثنين من التنبؤات. وبتطبيق هذه الأفكار العامة لحالة المتغيرات المقاسة على المستوى الاسمي سيجعل هذا المنطق أكثر وضوحاً، فالمتغيرات الاسمية، تقوم أولاً بالتنبؤ بالفئة التي سوف تقع على المتغير التابع (y) بينما تتجاهل المتغير المستقل (X). ولما كنا نتنبأ على نحو أعمى في هذه الحالة، فإننا سنرتكب أخطاء كثيرة (هذا يعني أنه في الغالب سنتنبأ بقيمة حالة على المتغير التابع بشكل غير صحيح).

إن التنبؤ الثاني يسمح لنا أن نأخذ المتغير المستقل في الاعتبار. فإذا كان هناك تطابق بين المتغيرين، فإن المعلومات التي يوفرها المتغير المستقل سوف تخفض أخطاء التنبؤ. وكلما كان التطابق بين المتغيرين قوياً، كان هناك تخفيض كبير في الأخطاء. وفي حالة التطابق التام لن تكون هناك أخطاء على الإطلاق عندما نقوم بعملية التنبؤ بدرجة y من درجة على X. وعندما لا يوجد تطابق بين المتغيرين، على الجانب الآخر، فإن المعرفة بالمتغير المستقل لن تحسن من دقة تنبؤاتنا. فإننا سوف نولد تماماً أخطاء كثيرة في التنبؤ بمعرفتنا بالمتغير المستقل مثلما فعلنا بدون معرفة المتغير المستقل.

والمثال التالي سوف يجعل هذه المبادئ أكثر وضوحاً. افترض أنك وجدت في موقف مألوف للتنبؤ بما إذا كان كل واحد من 100 شخص الذينَ قد قابلتهم سيكونون طوالاً أو قصاراً أكثر من 5 قدم و9 أنش (5.9) في القامة تحت ظرف أنك لا تمتلك أية معرفة حول هؤلاء الناس على الإطلاق. ولما كانت لا تتوفر أية معلومات حول هؤلاء الناس، فإن التنبؤات ستكون في الغالب خاطئة.

افترض الآن أنك مررت بهذه المحنة مرتين: ولكن في الجولة الثانية كان لديك معلومات حول جنس الشخص الذي تتنبأ بقامته. ولما كانت القامة مرتبطة بالنوع، وأن الإناث في المتوسط أقصر من الذكور، فإن الإستراتيجية الأفضل ستكون بأن التنبؤ سيكون أن كل الإناث قصيرات، وأن كل الذكور طوال. إلا أنه يمكننا القول، بأننا لازلنا نولد أخطاء في الجولة الثانية. ولكن إ ذا كان المتغيران مرتبطين فإن عدد الأخطاء سيكون أقل. بمعنى أنه إذا تم استخدام المعلومات حول المتغير المستقل سوف يخفض عدد الأخطاء.

فإذا رجعنا إلى متغير الدخل ومكان الإقامة ووجدنا أنهما قادران على التكهن حول أين يعيش شخص ما من خلال معرفتنا بمستوى دخله، وكلما كانت العلاقة قوية كان توقعنا صحيحاً.

إن كل مقاييس نسبة التخفيض في الخطأ تتبع نفس الإجراء. إننا نحاول أن نتنبأ كيف أن الحالات ستتوزع في الجدول الثنائي تحت شرطين:

1- يمكننا التكهن بتوزيع الحالات على طول المتغير التابع بدون أية معرفة في درجات هذه الحالات للمتغير المستقل.

2- يمكننا التكهن بتوزيع الحالات على طول المتغير التابع بمعرفتنا لدرجات هذه الحالات للمتغير المستقل، ولكي نبين كيف يمكننا أن نقوم بهذه التكهنات دعنا نفترض أن 720 شخصاً في المسح الذي نقوم به قد اصْطَفُّوا جميعاً خارج الحجرة ثم بدأ كل واحد بالتحرك تلو الآخر وقبل دخول أي شخص إلى الحجرة ينبغي علينا التخمين (التنبؤ) بما إذا كان هؤلاء يعيشون في مناطق ريفية أو حضرية (التنبؤ بدرجاتهم على المتغير التابع). ولفعل مثل هذه التوقعات فإننا تحصلنا فقط على جزء يسير من المعلومات التي تشير إلى أن الغالبية من كل الـِ 720 شخصاً يعيشون في مناطق حضرية.

إن السؤال الذي يمكن طرحه هنا: ما هو التخمين الذي يمكننا فعله قبل دخول أي شخص إلى الحجرة؟ إن معرفتنا فقط بأن الغالبية من الناس يعيشون في مناطق حضرية

هي أفضل تخمين للتكهن بأنّ كل 720 شخصاً يعيشون في مناطق حضرية. بمعنى آخر، عندما لا تكون لدينا معلومات أخرى، فإنه يمكن تخمين المتوسط! إن هذا في حقيقة الأمر هو استخدام نموذج لا تطابق الذي بيناه في جدول (4). كقاعدة للتنبؤ. فإذا لم تكن هناك علاقات كبيرة بين هذين المتغيرين فإن قاعدة التنبؤ سوف تولد لنا أخطاء قليلة جداً.

وعليه فإنه كلما كنا قريبين من أنماط الحالات الفعلية التي تشبه نموذج لا تطابق فقليل جداً من الأخطاء يمكن عملها عند استخدام هذه القاعدة للتنبؤ أين يعيش شخص ما؟ ففي المثال الذي بين أيدينا إذا ما توقعنا أن كل 720 شخصاً يعيشون في المناطق الحضرية، فإننا قد نعمل خطأ في التنبؤ يصل إلى 323 هذا هو الرقم لعدد الناس من المناطق الريفية التي لم نفلح في التكهن بأنهم يعيشون في مناطق حضرية، ونطلق على هذا الخطأ خطأ رقم (1) (E_1). 323 = E_1 .

دعنا الآن نفترض أن نفس 720 شخصاً تم إخراجهم من الحجرة وطُلب منهم أن يدخلوا مرة ثانية إلى الحجرة بشكل عشوائي واحد تلو الآخر. وسنجد في هذه المرة أنه قبل دخول أي شخص قد تم إخبارنا فيما إذا كان هؤلاء الناس لديهم دخل منخفض أو عالٍ وفي شكنا أنه لا يوجد تطابق بين الدخل ومكان الإقامة فإن مثل أولئك الناس ذوي الدخول المنخفضة يعيشون في مناطق ريفية وأن الناس ذوي الدخول العالية يعيشون في مناطق حضرية فإنه بإمكاننا التكهن بأن الشخص ذا الدخل المنخفض يعيش في الريف وأن الشخص ذا الدخل العالي يعيش في المدينة. إننا في هذه الحالة وبشكل فعال مستخدمون نموذج التطابق التام كما أشرنا إليه في الجدول (3).

وباتباعنا لقاعدة التكهن فإننا سنولد 285 خطأً. هذه الأخطاء جاءت من 167 لأولئك الذين يكسبون دخولاً منخفضة من المناطق الحضرية الذين تم تصنيفهم خطأ على أنهم يعيشون في الريف، هذا بالإضافة إلى 118 منخفضاً يكسبون دخلاً عالياً من المناطق الريفية الذين تم التكهن خطأً بأنهم يعيشون في المدن. ونطلق على مثل هذا الخطأ خطأ (2) (E_2). $E_2 = 167 + 118 = 285$.

إن السؤال الذي يمكن طرحه هو ما إذا كنا قد ولّدنا أخطاء قليلة عندما أعطيت لنا

معلومات إضافية حول الدرجة لكل حالة على المتغير المستقل. هل شكنا حول احتمالية التطابق بين المتغيرين تخفض معدل الخطأ عند توليدنا لهذه التنبؤات أم لا؟ إن التقليل في الأخطاء يكون 38 = 285 - 323. إذاً نحن ولدنا 38 من الأخطاء القليلة باستخدامنا لقاعدة التنبؤ التطابق التام أكثر منه عند استخدامنا لقاعدة التنبؤ لعدم التطابق مشيرين إلى أنه توجد بعض العلاقة في البيانات.

من خلال لامبيدا يمكننا حساب هذا التخفيض في الأخطاء كنسبة لـ (E_1).

$$\lambda = \frac{E_1 - E_2}{E_1}$$

حيث إن: E_1 هي عدد الأخطاء بدون معلومات حول المتغير المستقل.

E_2 عدد الأخطاء بمعلومات حول المتغير المستقل وكعملية تناسب فإن معدل الخطأ قد تم خفضه:

$$\lambda = \frac{E_1 - E_2}{E_1} = \frac{323 - 285}{323}$$

$$= 0.12$$

وعليه، بعد حصولنا على المعلومات حول مستوى دخل شخص ما (المتغير المستقل) فإنه بإمكاننا أن نقلل من الأخطاء إلى الحد الأدنى عندما نتكهن بِـ(أين يعيش شخص ما بِـ 12 %)؟ إن هذه النتيجة تبين لنا الميزة الكبيرة لمقاييس التخفيض في الخطأ (PRE): إن هذه المقاييس تقيس شيئاً ذا معنى يغير في التخفيض من معدلات الخطأ، وعليه يمكننا أن نتحصل على تفسيرات محددة. وكما نرى في شكل رقم 2 أن النتيجة وضعت التوزيعات المشاهدة للبيانات أكثر قرباً من أقصى لا تطابق من وجود تطابق تام.

لا يوجد تطابق — تطابق تام

0 — 1

شكل (7-2)

ومن هنا نجد أن لامبيدا تعطينا طريقة قاطعة، إنه بالرغم من وجود بعض العلاقة بين هذين المتغيرين، إلا أنها ليست علاقة قوية جداً. فنحن عادةً ما نتحدث بأن التطابق بين المتغيرات إما أن يكون ضعيفاً أو متوسطاً أو قوياً. وعليه لا يوجد خط فاصل دقيق يحدد متى تكون قيم مقاييس نسبة تخفيض الخطأ (PRE) ضعيفة؟ ومتى تكون قوية؟ إلا أنه يمكننا الاسترشاد بالتالي:

جدول (7-6) تفسير قيم لامبيدا

المدى (±)	القوة النسبية
0.0 - 0.2	ضعيفة جداً، جديرة بالإهمال
0.2 - 0.4	ضعيفة، تطابق منخفض
0.4 - 0.7	تطابق متوسط
0.7 - 0.9	قوية، عالية، تشير إلى تطابق
0.9 - 1.0	عالية جداً، علاقة قوية جداً

3- خصائص لامبيدا [3]:

تمتلك مقاييس التطابق للامبيدا خصائص معينة، بعضها يكون مرغوباً، في حين أن بعضها الآخر لا يكون كذلك من حيث التطبيق. وسوف نصف باختصار الخصائص الرئيسية في التالي:

أ- قيمة لامبيدا تصل دائماً إلى 1 عندما تشير البيانات إلى وجود تطابق تام. وإذا ما نظرنا إلى الطريقة التي تبنى بها لامبيدا فإن لامبيدا سيكون لها خصائص مقبولة في حالة التطابق التام الذي يصل إلى 1. فإذا كان لدينا تطابق تام بين المتغيرين، فإن البيانات سوف تتطابق تماماً مع قاعدة التنبؤ الثانية التي ينتج عنها عدم وجود أخطاء. وعندما لا تكون هناك أخطاء من خلال المعلومات المتعلقة بالمتغير المستقل

$(E_2 = 0)$ حينئذٍ ستكون معادلة لامبيدا ببساطة:

$$\lambda = \frac{E_1}{E_1} = 1$$

ب- إن قيمة لامبيدا غالباً تساوي 0 عندما تظهر البيانات أنه لا يوجد تطابق. فإذا لم يكن هناك تطابق في البيانات، فإن النتيجة المشاهدة سوف تطابق تماماً "نموذج لا تطابق"، وباستطاعتنا التكهن باستخدام "نموذج لا تطابق" الذي يولد لنا أخطاء $(E_2 = 0)$. وهذه الأخطاء سوف تولد لنا قيمة 0 للامبيدا.

ج- في بعض الأحيان ستكون لامبيدا مساوية لِـ 0 عندما تُظهر البيانات بعض التطابق. مع أن لامبيدا دائماً سوف تكون مساوية لِـ 0 عندما لا يوجد تطابق. والعكس ليس دائماً بالضرورة صحيحاً، ففي بعض الأحيان عندما تكون لامبيدا مساوية لِـ 0 فإنه قد يكون حقاً هناك تطابق. وهذا يمثل القصور الأساسي في استخدام لامبيدا والذي سنشير إليه في موضعه في نهاية هذا الفصل.

د- تعتبر لامبيدا مقياساً للتطابق اللاامتماثل Asymmetric. وهذا يعني أن قيمة لامبيدا سوف تكون مختلفة اعتماداً على الطريقة التي تم بها التفكير حول المتغيرين أيهما متغير مستقل وأيهما متغير تابع. بمعنى آخر، إذا ما حاولنا في المثال السابق أن نتكهن بمستويات الدخول استناداً على أين يعيش الأفراد مفضلين ذلك على الطريقة الأخرى، فإن قيمة لامبيدا ستتغير. وعليه عند استخدامنا للامبيدا ينبغي علينا أن نكون واضحين حول طبيعة العلاقة التي نعتقد أنها تربط هذين المتغيرين معاً. إن هذه الطريقة تجعل من لامبيدا طريقة مفيدة عندما يكون لدينا أسباب قوية للاعتقاد أنه توجد طريقة واحدة للعلاقة بين المتغيرين تسير في اتجاه محدد.

ولتوضيح كل ذلك دعنا نقلب المتغيرات في المثال السابق وأن نتعامل مع مكان الإقامة كمتغير مستقل، ومستوى الدخل كمتغير تابع. إن هذا الأمر يتطلب منا بناء جدول توافق حيث نضع متغير مكان الإقامة عبر الأعمدة والدخل وفقاً للصفوف (جدول 7 – 7).

جدول (7ـ7) مكان الإقامة وفقا لمستوى الدخل

مكان الإقامة	الدخل		المجموع
	منخفض	عال	
حضري	167	205	372
ريفي	230	118	348
المجموع	**397**	**323**	**720**

وبدون سابق أية معرفة لتوزيع المتغير المستقل (مكان الإقامة) فإنه باستطاعتنا التكهن بأن كل الحالات التي تقع في المجموعة ذات المستوى الاقتصادي المنخفض، لذلك فإن سوء التصنيف لـِ 348 شخصاً الذين يكسبون في الحقيقة دخلاً اقتصادياً عالياً:

$$E_1 = 348$$

وإذا ما أردنا أن نتكهن بمستويات الدخول لدى الناس، إلا أننا في هذه المرة لدينا معلومات فيما إذا كان هؤلاء يعيشون في مناطق ريفية أو حضرية، عندئذ يكون التكهن بأن كل الناس من المناطق الحضرية لديهم دخول عالية، ومن هنا نكون قد وقعنا في 118 خطأ. عليه فإن المعدل الكلي للخطأ عندما استخدمنا نموذج التطابق التام هو:

$$E_2 = 167 + 118 = 285$$

وبالتالي تصبح قيمة لامبيدا باستخدام نموذج العلاقة:

$$\lambda = \frac{E_1 - E_2}{E_1} = \frac{348 - 285}{348}$$

$$= 0.18$$

بمعنى آخر فإن التطابق يعتبر - إلى حد ما - ضعيفاً، إلا أن هذا التطابق قد يمدنا بالإحساس بأننا قادرون على أن نتوقع مستوى دخل شخص ما استناداً إلى المكان الذي يعيش فيه. فالتطابق بين هذين المتغيرين يعتبر إلى حد ما أقوى عندما نظرنا إلى مستوى الدخل كمتغير تابع بدلاً من مكان الإقامة.

إجراءات توليد لامبيدا Lambda باستخدام SPSS:

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Analyze ⟸ descriptive statistics ⟸ Crosstabs

2- انقر على المتغير في القائمة التي سوف تشكل صفوف الجدول، والتي تكون في هذه الحالة Place of Residence..

3- انقر على ◄ التي تشير إلى المساحة المعنونة Row(S)، تقوم بلصق Place of Residence في القائمة المحددة للصفوف Row(S) .

4- انقر على Income Level.

5- انقر على ◄ التي تشير إلى المساحة المعنونة Column(s). تقوم بلصق Income Level في القائمة المحددة للأعمدة Column(s).

6- انقر على زر Statistics، يقوم بإعطائنا صندوق Statistics: Crosstabs وفي أعلى زاوية من الشمال يمكننا رؤية المساحة المعنونة Nominal Data هذه مقاييس التطابق المتوفرة عندما يكون على الأقل أحد المتغيرات مقاساً على المستوى الاسمي. في هذا المثال Place of residence قد تم قياسه على المستوى الاسمي.

7- اختر Lambda بالنقر على الصندوق الموجود أمامها وضع علامة T في الصندوق لتبين أنه قد تم اختيار لامبيدا.

8- انقر على Continue.

9- انقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

Directional Measures

	Value	Asymp std. Error[a]	Aprrox. T[b]	Aprrox. Sig
Nominal by Lambda Symmetric	.151	.047	3.061	.002
Place of residence Dependent	.118	.056	1.976	.048
Income Level Dependent	.181	.052	3.184	.001
Goodman Place of residence Dependent	.045	.015		.000 c
And Kraskal-tau Income Level Dependent	.045	.015		.000 c

a. Not assuming the null hypothesis.
b. using the asymptotic standard error assuming the null hypothesis.
c. Based on chi-sguare approximation

في هذه المخرجات تم استخدام عملية التماثل عندما لا يكون لدينا أي سبب لأن نشك في أن واحداً من هذين المتغيرين يكون معتمداً على الآخر، بقدر ما يكون هذان المتغيران - تبادلياً - معتمدَيْنَ على بعضها البعض، ففي هذا المثال قيمة التماثل تساوي 0.151 تقع في مكان ما بين القيم اللاامتماثلة 0.118 و 0.181. إن عملية استخدام اللاتماثل تكون لديها قيمتان احتماليتان استناداً على أي من المتغيرين يعتقد أنه متغير تابع.

كذلك تحتوي هذه المخرجات على قيمة متعلقة بمقياس اسمي. ولقياس التطابق ألا وهو جودمان وكرشكال تاو Goodman and Kruskal tau والذي يحتوي على قيمة صغيرة للتطابق إذا ما قورن بـِ لامبيدا Lambda في الأعمدة الأخرى من هذا المربع تحتوي على معلومات ليست ذات علاقة بهذا الفصل، وإنما ستكون هذه المعلومات ذات أهمية في مواضع أخرى من هذا الكتاب خاصة عند التعامل مع الاستدلال من العينة على المجتمع [4].

حدود استخدام لامبيدا [5]:

على الرغم من الطريقة السهلة لحساب لامبيدا إلا أن هناك مشكلة عادة ما تقابلنا عند استخدام لامبيدا. إن هذه المشاكل التي تناولناها عند الحديث عن خصائص لامبيدا يمكن أن يكون لديها قيمة 0 حتى ولو كانت العلاقة موجودة بين المتغيرين. (قد تكون واضحة بمجرد النظر إلى جدول التقاطع) إن سبب هذه المشكلة يكمن في أن البيانات تنحرف بشكل كبير على طول المتغير التابع (إن قيمة لامبيدا ستكون 0 عندما تكون فئة النموذج للمتغير التابع مساوية لكل الفئات المتعلقة بالمتغير المستقل).

ولكي نرى ذلك عملياً، سوف نقوم بتحليل البيانات الواردة في الجدول التالي (7 - 8):

في هذا المثال النظري قد تم طرح السؤال المتعلق بما إذا كانت دائرة التشغيل تقوم بجهد كبير للتخفيف من مشكلة البطالة.

جدول (7-8):

هل يجب على دائرة التشغيل أن تبذل جهداً أكبر للتخفيف من البطالة

يوافق	العمر		المجموع
	أقل من 45	45 فما فوق	
لا	110 % 18	168 % 42	278
نعم	490 % 82	232 % 58	722
المجموع	600	400	1000

وبالنظر إلى جدول التقاطع فإنه يمكننا القول إنه توجد علاقة محدودة. فكلما زادت نسبة الأفراد الذين تكون أعمارهم تحت 45 سنة يوافقون على السؤال المطروح حول سياسة التشغيل أكثر من أولئك الذين تكون أعمارهم 45 سنة فأكثر. وبوضوح، يمكننا أن نرى أن هناك بعض الاعتماد بين المتغيرين، والذي يمكننا أن نصفه بتعبيرات لفظية بالقول إن هناك علاقة جيدة بعض الشيء إلى متوسطة في القوة.

وعليه إذا أردنا أن نحاول تكميم هذه العلاقة باستخدام لامبيدا فإننا نحصل على تطابق متناسب مساوياً لـِ 0. (لاحظ الإجابة للمتغير التابع لكل الحالات 1000 هي "نعم" والتي هي أيضاً كذلك الإجابة لكل من الفئتين للمتغير المستقل: فغالبية الناس تحت الـِ 45 سنة أجابوا بنعم، وأن الغالبية من الناس ذوي الأعمار 45 سنة فما فوق قد أجابوا هم أيضاً بنعم).

إن هذا التوزيع المنحرف في إطار المتغير التابع سوف يولد لنا لامبيدا تساوي 0، حتى كان واضحاً لدينا من خلال العين المجردة أن هناك وجوداً لبعض التطابق بين المتغيرين. ولكي نبين ذلك فإننا نحتاج أولاً: إلى حساب عدد الأخطاء عند التكهن بدون معلومات حول المتغير المستقل (العمر)، إننا نتوقع أن كل حالات الـِ 1000 سوف تقع

في فئة "نعم"، حيث إن هذا الأمر سوف يقلل معدل الخطأ حيث يصل معدل الخطأ في هذه الحالة إلى 278 خطأ: $E_1 = 278$.

أما إذا كانت لدينا معلومات حول المتغير التابع فإننا سوف نبقى على نفس الأخطاء. فإذا ما نظرنا إلى إجابة المبحوثين الأولى أقل من 45 سنة فإننا نتوقع أن 600 حالة قد أجابت "بنعم" (110خطأ).

ثانياً: يمكننا التوقع أن 400 شخص الذين تبلغ أعمارهم 45 سنة فما فوق قد أجابوا "بنعم" (168 خطأ).

وهذا المجموع الذي يصل إلى 278 خطأ هو نفس التوقعات بدون معرفة عمر المبحوث: $E_2 = 278$. وأن قيمة لامبيدا ستكون:

$$\lambda = \frac{E_1 - E_2}{E_1} = \frac{278 - 278}{278}$$

$$= 0$$

ومن هنا نجد أن لامبيدا قد فشلت في الحصول على علاقة مشاهدة والتي كانت واضحة لنا بالعين المجردة. إن هذا الأمر قد يلقي بعض الضوء على قاعدة مهمة: متى كانت لامبيدا مساوية لـِ 0 ينبغي معاينة التوزيع التكراري النسبي لنقرر فيما إذا كانت هذه التكرارات تعكس في الواقع لا علاقة أو فيما إذا كانت هذه العلاقة جاءت نتيجة لانحراف التوزيع المتعلق بالمتغير التابع. إذا فحصنا نسب العمود تقودنا إلى الاستنتاج أن قيمة 0 للامبيدا هي ناتجة من توزيع منحرف (كما في هذه الحالة) توجد لدينا ثلاثة خيارات:

1- ينبغي على الباحث ألا يشغل باله بمقاييس التطابق وعليه أن يتقيد بجدول التطابق وما يحتويه من توزيعات نسبية، وأن يكون الأساس في نتيجته هو النظر إلى العلاقة في ذاتها فقط. ويتطلب هذا الأمر من الباحث أن يقوم ببعض الأحكام الذاتية، ولكن طالما أن جدول التقاطع يسمح للقراء أن يقيموا بأنفسهم، فإن ذلك لا يشكل أي مشكلة في بناء هذه المحاولة باستخدام التوزيعات النسبية كشواهد. إن هذه

التوزيعات في بعض الأحيان "تتحدث بنفسها". فحساب إحصاءات متقدمة أكثر قد يؤدي إلى اختفاء معلومات مهمة في كتلة من الأعداد المشكوك فيها.

2- ينبغي على الباحث حساب مقاييس أخرى للتطابق حيث توجد مجموعة أخرى من مقاييس التطابق للبيانات المقاسة على المستوى الاسمي والتي يمكن استخدامها إذا ما واجهت الباحث بعض المشكلات المتعلقة بلامبيدا مثل مقاييس جودمان - كروشكل تاو Goodman-Kruskal tau. إن هذا المقياس يشبه مقياس لامبيدا في كونه مقياس لا تماثل للتطابق حيث تتراوح درجاته بين 0 و 1 ولكنه في الجانب الآخر لا يشبه لامبيدا لكونه لا يستخدم إجابة النموذج للمتغير المستقل عند إجراء عملية التنبؤات، ولكنه على العكس يستخدم بدلاً من ذلك التوزيع التكراري للحالات عبر كل الفئات للمتغير المستقل. ولما كان هذا الأمر أقل حساسية لانحراف التوزيعات الهامشية من لامبيدا بالتالي فهو خيار ملائم عندما تسبب عملية الانحراف قيمة لامبيدا المساوية لصفر.

ومن المقاييس الأخرى للتطابق هو Cramer`s V. وهذا المقياس دائماً يولد قيمة أكبر من 0 عندما يكون هناك تطابق بين المتغيرين. إلا أنه في الواقع لا يمتلك تفسيرات بسيطة فيما يتعلق بنسبة التخفيض في الخطأ (PRE).

وعليه فإنه لا يمكن استخدامه لتقييم قوة العلاقة في أي جدول تقاطع. إلا أنه يمكن أن يكون مفيداً عند مقارنة قوة العلاقات الثنائية عبر مجموعة من الجداول. والمعادلة المستخدمة لـِ Cramer`s V هي:

$$V = \sqrt{\frac{X^2}{N(K-1)}}$$

حيث إنّ:

X^2 هي إحصاءات مربع كاي لجدول التقاطع.

K تشير إلى عدد الصفوف أو إلى عدد الأعمدة أيهما أصغر. إنّ Crame`s V هي إحدى الخيارات الموازية مع لامبيدا، للبيانات الاسمية.

3- إن الخيار الثالث يكمن في معايرة الجدول حتى يكون مجموع الصفوف كلها متساوية. إن مثل هذا الإجراء إجراء معقد إلى حد ما، فعملية المعايرة في الجدول تتضمن محاولة التقليل من التباين في البيانات التي جاءت من خلال التوزيع المنحرف في المتغير التابع، بينما لا يزال الاحتفاظ بالتباين عبر فئات المتغير المستقل. وفي تقرير يستخدم هذا الإجراء ينبغي أن يكون واضحاً أن لامبيدا لا يمكن حسابها ببيانات الصف، وبحساب لامبيدا فإن هوامش الصفوف قد غُيّرت ليصل مجموعها إلى 100. وعند التعامل مع لامبيدا فإننا نقنن هوامش الصف بحيث يصل مجموع كل صف إلى 100. وهذا الأمر يتطلب حساب نسب الصف، التي يمكن التعامل معها كما لو كانت هي الأعداد الحقيقية للحالات. هذا يعني، أننا قمنا بحساب نسبة مجموع "نعم" للمبحوثين الذين هم تحت الـ 45 سنة، وحساب نسبة أولئك الذين تصل أعمارهم إلى 45 سنة أو تزيد. ونقوم بعمل الشيء نفسه لإجابة أولئك الذين أجابوا "بلا". انظر جدول (7 – 9).

جدول (7-9):

هل يجب على دائرة التشغيل أن تبذل جهدا كبيرا للتخفيف من البطالة

يوافق	العمر		المجموع
	أقل من 45	45 فما فوق	
لا	$\frac{110}{278} \times 100 = \%40$	$\frac{168}{278} \times 100 = \%60$	% 100
نعم	$\frac{490}{722} \times 100 = \%68$	$\frac{232}{722} \times 100 = \%32$	% 100

وبعد ذلك يمكننا استخدام أعداد النسب كما لو حسبت كحالات فعلية كما في الجدول التالي:

جدول (7ـ 10):

هل يجب على دائرة التشغيل أن تبذل جهداً كبيرا في التقليل من البطالة

يوافق	العمر		المجموع
	أقل من 45	45 فأكثر	
لا	40	60	100
نعم	68	32	100

تذكر أن هذه النسب: 40 تمثل 40 % من 278 لمجموع إجابات "لا". وهكذا. إلا أننا قد تعاملنا مع هذه النسب كما لو كانت حالات فردية. وهذا يعني أن الحجم الكلي للعينة هو 200 بدلاً من 1000: الذين أجابوا "بنعم" يصل عددهم إلى 100 والذين أجابوا "بلا" يصل عددهم إلى 100. وباستخدامنا لهذه البيانات من جدول المعايرة (10)، يمكننا إعادة حساب لامبيدا بدون معرفة حول المتغير المستقل، فيمكننا تصنيف كل الإجابات الـ 200 إما "بنعم" أو "بلا"، وعليه يمكننا أن نتسبب في أخطاء تصل إلى 100.

$$E_1 = 100$$

أما إذا كانت لدينا معرفة حول المتغير المستقل فإنه يمكن أن نصل إلى التوقعات التالية، فإذا بدأنا بأولئك الأشخاص الذين تقل أعمارهم عن 45 سنة، يمكننا أن نتنبأ بأولئك الذين أجابوا "بنعم"، ولما كان هذا التنبؤ يعطينا أقل معدل في الخطأ (40 خطأ) ولأولئك الذين تصل أعمارهم 45 سنة فما فوق، يمكننا التكهن بأنّ كل هؤلاء قد أجابوا "بلا"، وعليه فإننا بذلك قد ولدنا 32 خطأ:

$$E_2 = 40 + 32 + 72$$

ومن هنا تكون قيمة لامبيدا مساوية لـِ:

$$\lambda = \frac{E_1 - E_2}{E_1} = \frac{100 - 72}{100}$$

$$= 0.28$$

وعندما نقرب هذه القيمة (0.28) في المائة (100)، فإن قيمة لامبيدا تشير إلى تطابق ضعيف إلى متوسط، فيما يتعلق بنسبة التخفيض في الخطأ (PRE). وعليه، فإنه يمكن تفسير لامبيدا بالإقرار بأن معرفتنا بالعمر تساعد قدرتنا على أن نتنبأ بالاتجاه حول الجهد المبذول من دائرة التشغيل للتخفيف من البطالة بنسبة 28 %.

أسئلة للمراجعة:

1- بين الفرق بين لا تماثل Asymmetric و متماثل Symmetric، وما هي أنسب هذه المقاييس للاستخدام في موقف يكون فيه متغيران تبادلياً معتمدان على بعضهما البعض Mutually dependent؟

2- ما هو الشيء المهم عندما تقوم بحساب لامبيدا لنقرر ما إذا كان أحد المتغيرين يرجح أن يكون معتمداً على المتغير الآخر، وإذا كان الأمر كذلك، حدد ما هو المتغير التابع وما هو المتغير المستقل؟

3- احسب لامبيدا من البيانات التالية، وبين قوة أي علاقة؟

أ:

تابع	مستقل		المجموع
	1	2	
1	30	60	90
2	45	50	95
مجـ	75	110	185

ب:

تابع	مستقل			المجموع
	1	2	3	
1	56	40	10	106
2	15	30	50	95
مجـ	71	70	60	201

ج:

تابع	مستقل			المجموع
	1	2	3	
1	70	40	10	106
2	50	45	38	133
3	43	30	14	87
مجـ	163	115	62	326

4- يرغب باحث في معرفة الاتجاه نحو التعليم المختلط. وقد تم بحث 3000 طالبٍ وطالبةٍ، وجاءت النتيجة كالتالي:

الاتجاه	النوع		المجموع
	ذكور	إناث	
أوافق	849	367	
لا أوافق	191	1593	

احسب قيمة لامبيدا وفسرها.

5- احسب قيمة Cramerś V للبيانات التالية:

أ- $X^2 = 3.5, N = 20$

Culmns = 2 (الأعمدة)، rows = 4 (الصفوف)

Rows = 2, Culmns, $X^2 = 9.8, N = 90$ = 4

Rows = 3, Culmns, $X^2 = 12, N = 800$ = 3

6- مسح لعدد 50 طالبةً من الإناث و50 طالباً من الذكور، طرح عليهم سؤال هل تعرف تاريخ وضع حجر الأساس لمشروع النهر الصناعي العظيم؟

الذكور أجابوا "بنعم" 29

الإناث أجابوا "بلا" 21

الإناث أجابوا "بنعم" 22

الذكور أجابوا "بلا" 28

- رتب هذه البيانات في جدول تقاطع Crosstabs. وبين المتغير المستقل والمتغير التابع.
- احسب قيمة لامبيدا لهذه البيانات؟

7- دراسة توصلت إلى نتيجة مفادها أن التطابق بين متغيرين باستخدام Cramerś V، 0.34. وفي دراسة سابقة قامت بقياس التطابق بين نفس هذين المتغيرين باستخدام V كانت النتيجة تتراوح بين 0.15 إلى 0.21. كيف لهؤلاء الباحثين تفسير هذه النتائج؟

الهوامش والمصادر:

أولا: الهوامش:

1- George Argyrous, Statistics for Social and Health Research With a Guide to SPSS, Sage Publications ,London, 2001, P.153.

2- Ibid, PP. 154 - 155.

3- Ibid, PP. 159 - 161.

4- Ibid, P. 163.

5- George Argyrous, opt. Cit., PP. 165 - 196.

6- And, Joseph F. Healey, The Essentials of Statistics: A Tool for Social Research, Wadsworth Cengage Learning, USA, 2010. pp. 295 .

ثانيا: المصادر:

1- George Argyrous, Statistics for Social and Health Research With Guide to SPSS, Sage Publications, London, 2001 .

2- Joseph F. Healey, The Essentials of Statistitics: A Tool for Social Research, Wadsworth Cengate Learning, USA. 2010.

3- William E. Wagner, III, using SPSS for Social Statistics and Research Methods 2nd ed, Pine forge Press An Inprint of Sage, USA, 2010.

الفصل الثامن

التطابق بين متغيرات تم قياسها على المستوى الترتيبي

الوصف العددي للبيانات الترتيبية

مقدمة

نحاول في هذا الفصل التركيز على مقاييس التطابق لمتغيرات تم قياسها على المستوى الترتيبي لنسبة التخفيض من الخطأ (PRE) التي هي مشابهة لمقياس لامبيدا من حيث المنطق الأساسي لهذه المقاييس والكيفية التي يتم بها تفسير البيانات. وعند حديثنا عن لامبيدا فإننا نحاول من خلالها التنبؤ بقيمة حالة فردية أخذت للمتغير المستقل. والهدف من وراء ذلك هو أولاً: افتراض أنه لا يوجد تطابق بين المتغيرين. وثانياً: بافتراض أنه يوجد تطابق تام بين المتغيرين. وبمقارنة معدلات الخطأ المتعلق بكل قاعدة من قواعد التنبؤ فإننا نستطيع تقييم العلاقة الفعلية التي تحتويها مجموعة البيانات التي تم جمعها.

وتجدر الإشارة إلى أننا سنتبع نفس الإجراء مع البيانات الترتيبية، مع الأخذ في الاعتبار أننا سوف نستخدم بيانات إضافية حول المتغيرات المعطاة للتعامل معها وفقاً لمستوى البيانات الترتيبية، تختلف البيانات الترتيبية عن البيانات الاسمية، غير أننا في البيانات الترتيبية على دراية بالكيفية التي رتبت بها الحالات. وعليه فإننا نحاول أن نتنبأ

بالوضع الترتيبي لحالات الأزواج Pairs of Cases، وقاعدة مقاييس التطابق على أساس النجاح في تنبئنا لهذه الرتب. فإذا ما تعاملنا مع المثال الموضح في جدول (1) فإنه بإمكاننا أن نستمر في تكميم العلاقة القوية الموجبة ونلاحظ من خلال حساب مقاييس التطابق للبيانات الترتيبية ذات الصلة أنه توجد لدينا مجموعة من مقاييس التطابق المتعلقة بنسبة التخفيض في الخطأ التي يمكن حسابها من هذا الجدول مع تباين بسيط في طرقها الخاصة؛ وبذلك فإن كل المقاييس المتعلقة بالبيانات الترتيبية والتي سوف نناقشها لديها خصائص مشتركة تستند على التميِّز بين الأزواج المتوافقة Concordant Pairs والأزواج غير المتوافقة Disconcordant Pairs.

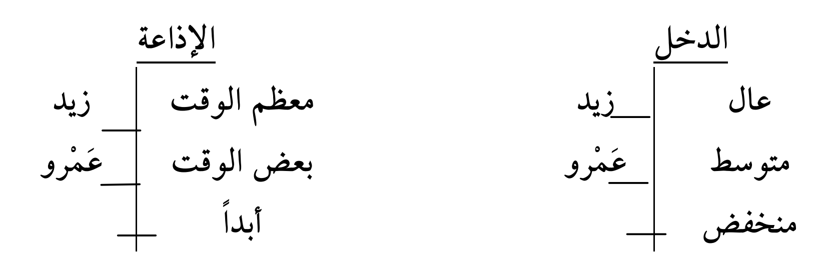
جدول (8ـ1): توزيع مشاهدة الإذاعة المرئية وفقا للدخل

مشاهدة الإذاعة المرئية	الدخل			المجموع
	منخفض	متوسط	عال	
أبداً	75 %71.5	15 %15.0	10 %10.5	100
بعض الوقت	20 %19.0	70 %70.0	10 %10.5	100
معظم الوقت	10 %9.5	15 %15.0	75 %79.0	100
المجموع	**105**	**100**	**95**	**300**

المصدر: George Argyrous, op.cit, p. 173

1- الأزواج المتوافقة:

نفترض أن واحداً من 75 في فئة الدخل العالي في جدول (1) يشاهد الإذاعة المرئية معظم الوقت ويدعى زيداً، وأن واحد من 70 في فئة الدخل المتوسط يشاهد الإذاعة المرئية بعض الوقت ويدعى عَمْراً. إن هذين الشخصين يمكن ترتيبهما بحيث يخالف أحدهما الآخر على كل واحد من هذين المتغيرين. انظر شكل (8 – 1).



شكل (8-1) ترتيب الزوج المتوافق

إن ترتيب زوج هاتين الحالتين يمكن تلخيصه في الجدول التالي:

جدول (8-2): حالات الزوجين المتوافقين

المتغير المستقل: الدخل	المتغير التابع: الإذاعة المرئية
زيد رتب فوق عَمْرو (لديه أعلى دخل)	زيد رتب أعلى من عَمْرو (مشاهدة أكثر للإذاعة المرئية)

وعليه فإن هاتين الحالتين قد تم ترتيبهما بشكل متساوٍ على كل واحد من هذين المتغيرين. وهذا قد يبدو غريباً في طرحه: حيث يمكن طرح السؤال التالي وهو: كيف يمكن أن يكون ترتيبهما متساوٍ إذا كان لكل منهما قيم مختلفة؟ إن القضية هي أن ترتيبهما متساوياً: إن زيداً قد رتب فوق عَمْرو على كلا المتغيرين. ونصف أزواج مثل هذه الحالات كزوج متوافق (Nc). ونعني بالزوج المتوافق: هو ذلك الزوج الذي يتشكل من خلال حالتين في توزيع متحد الذي رتب بشكل متساوٍ على كلا المتغيرين.

لقد أخذنا حالتين من كل الحالات البالغ عددها 300 والتي تشكل الزوج المتوافق. إن السؤال المطروح هو: هل نقوم بحساب العدد الكلي للأزواج المتوافقة التي يحتويها الجدول؟ ولفعل ذلك انظر إلى الخلايا المظَّلَّة في جدول التوافق التي من خلالها تم سحب كل من زيد وعَمْرو.

جدول (8-3) توزيع مشاهدة الإذاعة المرئية وفقا للدخل

مشاهدة الإذاعة المرئية	الدخل		
	منخفض	متوسط	عال
أبداً	75	15	10
بعض الوقت	20	70	10
معظم الوقت	10	15	75

في مناقشتنا أعلاه قد تم صياغة الزوج المتوافق بمماثلة زيد الذي هو واحد من الحالات التي تبلغ 75 حالة ولديه دخل عالٍ ويشاهد الإذاعة المرئية معظم الوقت. بعَمْرو الذي هو واحد من 70 حالة ذات الدخول المتوسطة ويشاهد الإذاعة المرئية بعض الوقت.

في واقع الأمر يمكننا أن نُقْرِنَ عَمْراً مع كل واحد وأي واحد من 70 حالة في الخلية التي تحتوي على دخل متوسط / ومشاهدة الإذاعة بعض الوقت يولد لنا 70 حالة من الأزواج المتوافقة: عَمْرو مضافاً إلى كل حالة من السبعين حالة الموجودة في الخلية الوسطى من الجدول (شاملة عَمْراً). وعندئذٍ يمكننا عمل الشيء نفسه لكل واحد من 74 حالة التي لديها دخل عالٍ وتشاهد الإذاعة المرئية معظم الوقت. إن هذا سوف يعطينا في المجموع النهائي 75 حصة في 70 أزواج متوافقة: $5250 = 70 \times 75$.

وبالنظر إلى جدول (8 – 4) يمكننا أن نرى أن هناك مجموعات متحدة التي بدورها تشكل الأزواج المتوافقة.

جدول (8-4): توزيع مشاهدة الإذاعة المرئية وفقاً للدخل

مشاهدة الإذاعة المرئية	الدخل		
	منخفض	متوسط	عال
أبداً	75	15	10
بعض الوقت	20	70	10
معظم الوقت	10	15	75

إن كل واحد من 75 حالة في أدنى الخلية اليسرى أيضاً تم ترتيبه فوق كل حالة من الحالات 15 في خلية أبداً / الدخل المتوسط كلاهما لديه دخل عالٍ ويشاهد الإذاعة المرئية أكثر. وبهذا سوف نضيف الأرقام التالية للأزواج المتوافقة:

$$75 \times 15 = 1125$$

في الحقيقة أن أي حالة سوف تشكل زوجاً متوافقاً مع أي حالة أخرى في الخلية التي هي فوق وإلى اليسار منها في الجدول وبالتالي فإن العدد الإجمالي للأزواج المتوافقة سيكون مبيناً في الجدول (8 – 5).

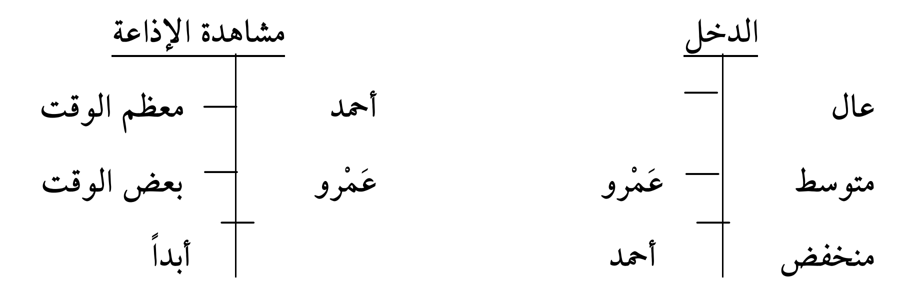
جدول (8-5): حساب الأزواج المتوافقة

75	15	10	
20	70	10	(75×70) +75×15) +75×20) + (75×75 = 13.500
10	15	75	
75	15	10	
20	70	10	(10 × 15) + (10 × 75) = 900
10	15	75	
75	15	10	
20	70	10	(15 × 20) + (15 × 75) = 1425
10	15	75	
75	15	10	
20	70	10	(70 × 75) = 5250
10	15	75	

$$N_C = 13.500 + 900 + 1425 + 5250 = 21.075$$

2- الأزواج غير المتوافقة:

دعنا الآن نأخذ واحداً من عشرة أشخاص الذين لديهم دخل منخفض ويشاهدون الإذاعة المرئية معظم الوقت، ويدعى أحمد، ومقارنته بعَمْرٍو (هو واحد من 70 حالة ذوي الدخول المتوسطة ويشاهدون الإذاعة المرئية بعض الوقت)، فإن الترتيب لن يكون متساوياً لكلا المتغيرين فأحمد تم ترتيبه تحت عَمْرٍو فيما يتعلق بالدخل، ولكنه على الطرف الآخر رتب فوق عَمْرٍو فيما يتعلق بمشاهدة الإذاعة المرئية، انظر شكل (8 – 2). إن مثل هذه الحالات نطلق عليها الأزواج غير المتماثلة (Nd). ونعني بالزوج غير المتماثل هو ذلك الزوج الذي يتشكل من خلال حالتين في توزيع متحد الذي يرتب فيه متغير واحد بشكل مختلف في ترتيبه للمتغير الآخر.



شكل (8-2) ترتيب الزوج غير المتوافق

هذه الحالة سوف تشكل زوجاً غير متوافق مع الحالة الأخرى في الجدول أي في أي خلية أعلى وإلى اليمين. ولحساب العدد الإجمالي للأزواج غير المتوافقة ينبغي علينا أن نبدأ بالخلية اليسرى أسفل الجدول (8 – 6) وتماثلها مع كل الخلايا فوق وإلى اليمين منها.

جدول (8ـ6): حساب الأزواج غير المتوافقة

75	15	10	
20	70	10	$(10\times70) + (10\times15) + (10\times10) + (10\times10) = 1050$
10	15	75	
75	15	10	
20	70	10	$(20 \times 15) + (20 \times 10) = 500$
10	15	75	
75	15	10	
20	70	10	$(15 \times 10) + (15 \times 10) = 300$
10	15	75	
75	15	10	
20	70	10	$(70 \times 10) = 700$
10	15	75	

$$N_d = 1050 + 500 + 300 + 700 = 5220$$

مقاييس التطابق للمتغيرات الترتيبية:

كل نسب التخفيض في الخطأ في مقاييس التطابق (PRE) تستخدم الفرق بين الأزواج المتوافقة والأزواج غير المتوافقة كأساس لتقييم ما إذا كان هناك تطابق وتحديد اتجاه هذا التطابق. إن السبب وراء النظر حول هذه الأزواج المتوافقة وغير المتوافقة أنها تعطينا معلومات يمكننا استخدامها في التنبؤ بما إذا كان لدينا متغيران متطابقان بشكل

موجب بعدئذٍ فإن جدول التقاطع سوف يحتوي أكثر الأزواج المتوافقة منه من الأزواج غير المتوافقة، والعكس بالعكس عندما يكون لدينا تطابق سالب.

وتجدر الإشارة إلى أنه عندما يكون هناك تطابق موجب بين المتغيرين فإن البيانات سوف تحتوي كثيراً من الأزواج المتوافقة وقليلاً من الأزواج غير المتوافقة. وإذا كان الأمر كذلك، فإن علمنا بأن شخصاً ما ترتيبه فوق شخص آخر فيما يتعلق بالدخل عندئذٍ يمكننا أن نتوقع أن ذاك الشخص يحتل كذلك مرتبة أعلى فوق الشخص الآخر فيما يتعلق بتوزيع مشاهدة الإذاعة المرئية.

التطابق الموجب: $N_c - N_d > 0$

أما في حالة العلاقة السالبة بين المتغيرين فإن البيانات سوف تحتوي على كثير من الأزواج غير المتوافقة ومن هنا يمكننا الوصول إلى تكهن عكسي: فمعرفتنا بشخص رتب فوق شخص آخر فيما يتعلق بالدخل سيقودنا إلى أن نتكهن بأن هذا الشخص يحتل مرتبة أدنى من الشخص الآخر فيما يتعلق بتوزيعات مشاهدة الإذاعة المرئية:

التطابق السالب: $N_c - N_d < 0$

أما فيما يتعلق بعدم وجود تطابق بين المتغيرين فإن البيانات ستحتوي تماماً على كثير من الأزواج المتوافقة والأزواج غير المتوافقة ومن هنا نكون غير قادرين على زيادة قدرتنا في التكهن بفئة المتغير التابع من خلال معرفتنا بفئة المتغير المستقل التي تقع فيها هذه الفئة:

لا تطابق: $N_c - N_d = 0$

وتجدر الإشارة هنا إلى أن هناك أربعة أسس لنسبة التخفيض في الخطأ في مقاييس التطابق للبيانات الترتيبية: سومرز Somerś d d وجاما Gamma وكنلدتاو Kendalś tau-b b وكنلدتاو Kendalś tau-c C. كل هذه المقاييس متشابهة في كونها تعطينا تفسيرات لنسبة التخفيض في الخطأ، كما أن كل هذه المقاييس تستخدم الفرق بين Nc و Nd كأساس لتقييم قوة العلاقة. إلا أن الفرق بين هذه المقاييس يكمن في كيف يمكننا معايرة هذا الفرق. ولمناقشة هذه المقاييس دعنا نبدأ بأبسطها المتمثل في جاما.

جاما Gamma:

تعتبر جاما من المقاييس الشائعة لـِ (PRE) لقياس التطابق بين متغيرين تم قياسهما على الأقل على المستوى الترتيبي وتم ترتيب هذين المتغيرين في جدول ثنائي. كما تعتبر جاما مقياساً متماثلاً للتطابق. وعليه فإن القيمة المحسوبة ستكون واحدة بغض النظر عن الكيفية التي يتحدد فيها المتغير المستقل والطريقة التي يحدد بها المتغير التابع. بمعنى آخر، إذا قلبنا الصفوف والأعمدة في الجدول بمعنى أن يكون الدخل تحت عند الصفوف ومشاهدة الإذاعة المرئية يكون عبر الأعمدة. فإن حساب جاما لن يتأثر. وعليه فإن جاما لن تكون حساسة لنموذج معين نعتقد أنه يصف العلاقة بين المتغيرين.

إن معادلة جاما تبين لنا الفرق بين عدد الأزواج المتوافقة والأزواج غير المتوافقة كنسبة من العدد الكلي للأزواج المتوافقة والأزواج غير المتوافقة وباستخدام البيانات من المثال السابق فإن حساب جاما يكون كالتالي:

$$G = \frac{N_C - N_d}{N_C + N_d} = \frac{21.075 - 2550}{21.075 + 2550}$$

$$= 0.78$$

وتشير هذه النتيجة إلى تطابق قوي موجب بين هذين المتغيرين، تلك النتيجة التي تعزز النتيجة التي وصلنا إليها استناداً إلى التحليل النظري لجدول التطابق. إنه من الوضوح أن المدى الممكن لقيم جاما تتراوح بين 1- و +1. فعندما تكون جاما 1- فإنها تشير إلى تطابق تام سالب: فمعرفة أن الحالة مرتبة فوق الحالة الأخرى على واحد من المتغيرين تشير إلى أنه يجب أن يرتب تحت المتغير الآخر. إن مثل هذه النتيجة نتحصل عليها إذا كان لدينا فقط الأزواج غير المتوافقة كما يوضحه الجدول التالي:

جدول (8ـ7)

توزيعات مشاهدة الإذاعة المرئية وفقا للدخل: تطابق تام سالب

مشاهدة الإذاعة المرئية	الدخل		
	منخفض	متوسط	عال
أبداً	0	0	%100
بعض الأوقات	0	%100	0
كل الأوقات	%100	0	0

أما على الجانب الآخر، إذا كان لدينا فقط أزواج متماثلة فإن قيمة جاما ستكون +1 مشيرة إلى تطابق موجب تام يعني: معرفتنا بترتيب حالة فوق الحالة الأخرى على المتغير المستقل تشير إلى أن هذه الحالة يجب أيضاً أن تكون مرتبة فوق المتغير التابع. إن مثل هذا الموقف ينعكس في الجدول التالي:

جدول (8ـ8)

توزيعات مشاهدة الإذاعة المرئية وفقا للدخل: تطابق تام موجب

مشاهدة الإذاعة المرئية	الدخل		
	منخفض	متوسط	عال
أبداً	%100	0	0
بعض الأوقات	0	%100	0
معظم الأوقات	0	0	%100

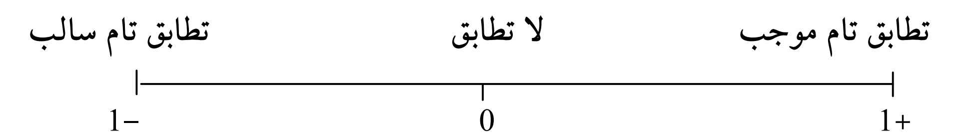
عندما تكون قيمة جاما صفر فهي تشير إلى عدم وجود تطابق. فإذا كان هناك فقط مجموعة كبيرة من الأزواج المتماثلة مثل وجود مجموعة كبيرة من الأزواج غير المتماثلة، عندئذٍ فإن معرفة الترتيب على طول متغير واحد لا يعطينا أي دليل في كيفية ترتيبه على المتغير الآخر. ويتضح هذا الموقف جلياً في الجدول التالي:

جدول (8-9)

توزيعات مشاهدة الإذاعة المرئية وفقا للدخل: لا تطابق

مشاهدة الإذاعة المرئية	الدخل		
	منخفض	متوسط	عال
أبداً	%50	0	%50
بعض الأوقات	0	%100	0
معظم الأوقات	%50	0	%50

تشير الجداول السابقة إلى توضيح النقاط الثلاثة المتطرفة على المقياس المعياري لقياس قوة التطابق بين متغيرين ترتيبيين. كما هو موضح في الشكل التالي:



شكل (8-3) مدى قيمة جاما

وبشكل واضح فإن البيانات الواردة في هذا المثال لا تتطابق مع أي واحد من هذه المواقف الثلاثة المتطرفة. إذاً السؤال الذي يطرح الآن هو أي من قواعد التنبؤ ستكون قريبة من النتائج التي تحصلنا عليها فعلياً؟ إنه من الوضوح أن جدول التطابق التام الموجب يمثل واحدة من تلك البيانات الفعلية الأقرب تشابهاً. فقيم جاما التي تحصلنا عليها (0.78) لا نجدها تماماً 1+ ولكنها قريبة منه أكثر منه إلى 0 أو –1.

وتمتاز جاما بشيوعية استخدامها نتيجة لسهولة حسابها نسبياً، إلا أن هذه الميزة قد أبطلت باستخدام برنامج العقل الآلي مثل (SPSS) الذي من خلاله يمكن حساب كل المقاييس بكل سهولة ويسر. إلا أنه قد يساورنا الشك أن عنصراً آخر قد يكون مرتبطاً بشيوعية استخدام جاما ولديه القدرة على توليد قيمة عالية لقوة التطابق مقارنة بمقاييس

تطابق أخرى ترتيبية. وتجدر الإشارة إلى القول بأن جاما تعتريها بعض القيود ينبغي علينا أن نكون على وعي بها:

أولهما: أن مقياس جاما هو فقط مقياس متماثل، وعليه لا يأخذ ميزة المعلومات المقدمة من خلال البيانات التي نعتقد أنها نموذج أكثر ملاءمةً لوصف العلاقة التي تكون معتمدة على طريقة واحدة ويتمثل هذا القيد الرئيسي الآخر عندما يكون التطابق تاماً سوف تولد لنا جاما قيمة +1 أو −1 والعكس ليس دائماً صحيحاً: فقيمة جاما + أو −1 لا تشير دائماً إلى تطابق تام. فقد يكون بإمكاننا أن نولد قيمة لجاما −1 أو +1 لجدول تقاطع حتى عندما يكون هناك بشكل واضح أقل من التطابق التام في البيانات. ويحدث هذا الأمر عندما تكون العلاقة غير مُتسقة. هنا ينبغي علينا اتباع القاعدة، وعليه فإنه قبل استخدام جاما لبيانات جدول ثنائي ينبغي علينا فحصها لتقييم ما إذا كانت العلاقة علاقة مُتسقة. كلا هذين القيدين في حقيقة الأمر ناشئُ من نفس خاصية حساب جاما. وهذا يتضمن عجزاً مرتبطاً بجاما يتضمن الحالات المتعادلة في معادلتهما. وعليه توجد ثلاثة أنماط من الحالات المتعادلة:

1- الحالات المتعادلة على المتغير المستقل (Tx) هذه الأزواج للحالات التي لديها نفس القيمة للمتغير المستقل ولكنها في ذات الوقت لديها قيم مختلفة للمتغير التابع. وهذا في الواقع أن أي حالتين في نفس عمود جدول التقاطع ولكنهما في صفوف مختلفة. في المثال السابق هذه الأزواج من الحالات التي لديها نفس الدخل ولكنها تشاهد كمية مختلفة من الإذاعة المرئية.

2- الحالات المتعادلة على المتغير التابع (Ty) فإن هذه الأزواج للحالات لديها نفس القيمة للمتغير التابع ولكنها في ذات الوقت لديها قيم مختلفة للمتغير المستقل. ومن الناحية العملية، فإن أي حالتين في نفس الصف بجدول التقاطع ولكنها في أعمدة مختلفة. في المثال الذي بين أيدينا هذه الأزواج من الحالات تشاهد نفس كمية الوقت ولكن لديها دخل منخفض.

3- حالات متعادلة على كلا المتغيرين (Tyx) هذه الحالات لديها نفس القيمة في كلا المتغيرين. فهذه أزواج من الحالات سحبت من نفس الخلية في الجدول. ففي المثال

الذي أوردناه فإن هذه الأزواج من الحالات لديها نفس الدخل وتشاهد نفس الكمية من الوقت "الإذاعة المرئية".

وأخيراً توجد مقاييس أخرى متعلقة بنسبة التخفيض في الخطأ في مقاييس التطابق للبيانات الترتيبية التي نسعى إلى تعويض هذا القصور المرتبط بجاما وذلك بإضافة بعض أو كل هذه الحالات المتعادلة في عملية حساباتها [2].

سومرز Somerś d d:

يعتبر مقياس سومرز d مقياساً غير متماثل للتطابق وبذلك يكون هذا المقياس ذا حساسية لأي متغير يوصف بأنه متغير مستقل أو يوصف آخر بأنه متغير تابع. وعليه فإنه من المفيد عندما نشعر بالعلاقة بين متغيرين أن أفضل وصف لهذه العلاقة يتوقف على نموذج اعتماد الاتجاه الواحد. إن المنطق وراء استخدام سومرز d يعتمد في الأساس على فكرة أن الحالتين اللتين تتباينان فيما يتعلق بالمتغير المستقل ولكنهما لا تتباينان فيما يتعلق بالمتغير التابع (هذه الحالات متعادلة على المتغير التابع) تعكس لا تطابق. في المثال الذي تعاملنا معه، فالأزواج المتعادلة على المتغير التابع ولكنها ليست كذلك على المتغير المستقل فإن تلك الأزواج من الحالات تختلف فيما يتعلق بالدخل ولكن مشاهدة التطابق تماماً نفس كمية الإذاعة المرئية. وتحسب قيمة سومرز d كنسبة كل الحالات المتوافقة وغير المتوافقة مضافاً إليها الأزواج المتعادلة على المتغير التابع:

$$d = \frac{N_C - N_d}{N_C + N_d + T_y}$$

ولحساب عدد الحالات المتعادلة فإننا نأخذ كل خلية، بَدْءًا من أعلى اليسار ونضربها في عدد الحالات التي تحتويها في عدد الحالات في الخلايا إلى اليمين منها. انظر جدول (8 – 10) وبتعويض هذه الحسابات في معادلة سومرز d نحصل على القيمة التالية:

$$d = \frac{N_C - N_d}{N_C + N_d + T_y} = \frac{21.075 - 2550}{21.075 + 2550 + 6350}$$

$$= 0.62$$

إن هذه القيمة 0.62 تشير إلى تطابق موجب، متوسط بين هذين المتغيرين فالزيادة في الدخل مرتبطة بالزيادة في مشاهدة الإذاعة المرئية.

جدول (8-10) حساب الحالات المتعادلة على المتغير التابع

75	15	10	
20	70	10	$(75\times15) + (75\times10) = 1875$
10	15	75	
75	15	10	
20	70	10	$(15 \times 10) = 150$
10	15	75	
75	15	10	
20	70	10	$(20 \times 70) + (20 \times 10) = 1600$
10	15	75	
75	15	10	
20	70	10	$(70 \times 10) = 700$
10	15	75	
75	15	10	
20	70	10	$(10 \times 15) + (10 \times 75) = 900$
10	15	75	
75	15	10	
20	70	10	$(15 \times 75) = 1125$
10	15	75	

$$T_y = 1875 + 150 + 1600 + 700 + 900 + 1125 = 3506$$

لاحظ أن معادلة سومرز d تقريباً تكون مساوية لمعادلة جاما ما عدا أن المقام في معادلة سومرز d لإعداد المتغير التابع تكون متعادلة. وكنتيجة لذلك عندما يكون لدينا مثل هذه الحالات المتعادلة، فإن قيمة سومرز d ستكون دائماً قيمتها أقل من قيمة جاما.

بمعنى آخر، بتجاهل الحالات المتعادلة فإن جاما يمكن أن يبالغ في قوة التطابق بين المتغيرين في علاقة غير متماثلة، لاسيما عندما تكون هناك مجموعة كبيرة من الحالات المتعادلة وحيث إنّ سومرز d هو مقياس غير متماثل للتطابق فإنه بإمكاننا في الواقع حساب خيارين لهذا المقياس، لأي جدول تقاطع. بمعنى آخر، يمكننا حساب سومرز d من خلال متغير واحد كمتغير مستقل والمتغير الآخر كمتغير تابع، كما أنه بإمكاننا قلب هذين المتغيرين بكيفية أخرى وحساب قيمة سومرز d مرة ثانية. ففي المثال السابق قمنا بحساب سومرز d متعامِلِينَ مع الدخل كمتغير مستقل ومشاهدة الإذاعة المرئية كمتغير تابع، وبما أن النموذج النظري لهذه العلاقة يصور لنا أن السببية تسير في ذلك الاتجاه فقد يكون شخص ما لديه نظرية مختلفة التي تفترض بطريقة ما أن مشاهدة الإذاعة المرئية تحدد مستوى الدخل استناداً إلى ذلك يصبح في مقدوره حساب قيمة سومرز d كخيار آخر مستخدماً الدخل كمتغير تابع متحصلاً على قيمة مختلفة لـِ سومرز d.

كندل - تاو Kendallś tau-b: b:

يعتبر كندل تاو b مقياساً للتطابق غير متماثل لنسبة التخفيض في الخطأ (PRE) لبيانات ترتيبية تم ترتيبها في جدول ثنائي. والميزة الرئيسية المرتبطة بمقياس كندل تاو b تكمن في استخدامه للمعلومات التي توفرها الحالات المتعادلة على المتغير التابع والمتغير المستقل:

ومن خلال المبادئ الرياضية يمكننا ملاحظة أن تاو b تمثل الوسط الهندسي للقيم البديلة لسومرز d. وبالرغم من أن هذا المفهوم قد يكون مشوشاً إلى حد ما نظراً لأن سومرز d من خلال التعريف هو مقياس غير متماثل. ولما كانت قيمة تاو b هي الوسط الهندسي لمقياس سومرز d فإن هذه القيمة ستكون في مكان ما بين القيمتين المتعلقتين بسومرز d التي يمكن حسابها من أي جدول تقاطع أمكن استخدامه. وتجدر الإشارة، إلى أن قيمة تاو b يتراوح مداها بين -1 و +1 حيث الأرقام المتعلقة بالصفوف تكون مساوية لعدد الأعمدة. وعليه وبشكل عام فإن تاو b تستخدم في حالات خاصة.

كندل تاو - C Kendallś tau-c:

يعتبر كنلدتاو C مقياساً متماثلاً لنسبة التخفيض في الخطأ للبيانات يشبه كثيراً مقياس تاو b. ويستخدم تاو C في مواقف تكون فيها رغبة الباحث في قياس التماثل في جدول يحتوي على أعداد غير متساوية من الصفوف والأعمدة التي تحتوي على عدد كبير من الحالات المتعادلة. إن المعادلة الدقيقة لِـ tau-c:

$$tau - C = \frac{2k(N_C - N_d)}{N^2(k-1)}$$

حيث إنّ:

K عدد الصفوف أو عدد الأعمدة أيهما أصغر،

N مجموع عدد الحالات.

في المثال السابق لدينا نفس عدد الصفوف مساوياً لعدد الأعمدة (ثلاثة) لذلك فإن

K = 3

وأن قيمة tau-c ستكون:

$$tau - C = \frac{2k(N_C - N_d)}{N^2(k-1)} = \frac{2(3)(21.075 - 2550)}{300^2(3-1)}$$

$$= 0.62$$

ويتضح مما سبق أن هذه النتيجة مساوية لقيمة سومرز d التي تم حسابها فيما سبق ولذلك فإن هذين البديلين من المقاييس تتفق مع بعضها البعض في النتائج [3].

إجراء مقاييس التطابق باستخدام SPSS:

1- من القائمة الموجودة في الجزء العلوي من الشاشة انقر فوق:

Crosstabs ⇐ Statistics ⇐ Descriptive ⇐ Analyze

2- انقر فوق Frequency of TV. Watching.

3- انقر فوق ◄ الذي يشير إلى Row(S) في القائمة المحددة للمتغيرات Variables List.

4- انقر على المتغير الذي يشكل الصفوف في الجدول، في هذه الحالة Income Level.

5- انقر على ◄الذي يشير إلى Column(s) في القائمة المحددة للمتغيرات Variable List ستقود هذه العملية إلى لصق Income Level في Column(s) في القائمة المحددة للمتغيرات Income Level.

6- انقر على المربع الذي يكتب عليه Statistics، فيظهر مربع Statistics box: Crosstabs، وهنا تجد مساحة معنونة بـِ: Ordinal التي توفر قائمة بمقاييس التطابق المتوفرة لهذا المستوى من القياس.

7- يختار الباحث معامل Gamma و Somerś dبالنقر على الصندوق التالي لهما أو القريب إليهما. ثم وضع علامة T على صندوق المعامل المطلوبة للإشارة إلى الإحصاءات المختارة.

8- انقر فوق Continue.

9- أنقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

تفسير مقاييس التطابق من مخرجات SPSS:

يمكن للباحث أن يتحصل على نتائج مقاييس التطابق من مخرجات SPSS. وهذه المقاييس هي جزء من أوامر جدول التقاطع Crosstabs Command.

والشكل التالي يبين مخرجات SPSS لإحصاءات جداول التوافق Crosstabs Statistics.

Directional Measures

	Value	Asymp. Stand. Error[a]	Appro T_b	Appro sig.
Ordinal by Ordinal Sommerś d Symmetric Frequency of TV. Watching	.618	0.043	14.192	.000
Dependent	.618	0.043	14.192	.000
Income Level dependent	.618	0.043	14.192	.000

a. Not Assuming the Null hypothesis.

b. using the Asymptotic standard Error assuming the Null hypothesis.

Symmetric Measures

	Value	Asymp. Stand. Error[a]	Appro T_b	Appro sig.
Ordinal by Ordinal Gamma	.784	0.043	14.192	.000
N. of Valid Cases	300			

a. Not Assuming the Null hypothesis.

b. using the Asymptotic standard Error assuming the Null hypothesis

المصدر: George Argyrous, op.cit, P. 184

شكل (8-4): مخرجات SPSS لجداول التقاطع: مخرجات إحصائية

إن حزمة SPSS تنتج لنا القيم المتعلقة بـِ Gamma، وسومرز d: Sommerś d، في مربعين منفصلين يمثل أحدهما مقياساً متماثلاً والآخر مقياساً غير متماثل (يطلق عليه مقياس مباشر Directional Measures في حزمة SPSS. وفي كلتا الحالتين نجد أن القيمتين يولدهما برنامج SPSS، في العمود المعنون "بالقيمة" Value في كل مربع، فهي نفس القيم التي تم حسابها يدوياً. وتعكس هذه القيم تطابقاً متوسطاً، إلى تطابق قوي بين هذين المتغيرين لهذه الحالات المدروسة. وإذا وجد تطابق سالب، فإن الإشارة ناقص (−) سوف تطبع أمام القيمة.

معامل سبيرمان للرتب [4]:

كما نعرف أن ارتباط بيرسون (r) يقيس درجة العلاقة الخطية Linear relationship بين متغيرين عندما تكون البيانات (قيم X و y) تحتوي على درجات عددية مقاسة على المستوى ذي المسافات أو النسبي، وعلى أية حال، فقد تطورت مُعاملات أخرى للعلاقات غير الخطية NonLinear relationship لأنماط أخرى من البيانات، وأن أحد أهم هذه المقاييس قد أطلق عليه ارتباط سبيرمان. ويستخدم ارتباط سبيرمان في موقفين أولهما تستخدم معامل سبيرمان لقياس العلاقة بين X و y، عندما يكون هذان المتغيران مقاسين على مستوى المقياس الترتيبي، وثانيهما تعتبر معامل سبيرمان للرتب بديلاً ذا قيمة لمعامل بيرسون (r) حتى عندما تكون الدرجات الخام الأصلية مقاسة على المستوى ذي المسافات / والنسبي.

إذاً يستخدم معامل سبيرمان في موقفين هامين هما:

1- عندما تكون البيانات الأصلية بيانات ترتيبية أي عندما ترتب قيم X و y. في هذه الحالة، يمكننا ببساطة تطبيق معادلة ارتباط بيرسون (r) لمجموعة الرتب.

2- يستخدم معامل ارتباط سبيرمان عندما يرغب الباحث في مقياس اتساق العلاقة بين X و y. في هذه الحالة، تحول الدرجات الأصلية أولاً إلى رتب؛ وبعد ذلك تستخدم معادلة معامل بيرسون (r) للرتب، باعتبار أن معادلة بيرسون تقيس الدرجة على

النحو الذي تنطبق فيه الرتب على الخط المستقيم a straight Line، كما أن معادلة بيرسون (r) تقيس درجة الاتساق في العلاقة للدرجات الأصلية، وبالمصادفة، عندما يوجد اتساق في علاقة ذات اتجاه واحد بين المتغيرين فإننا نطلق على هذه العلاقة علاقة التماثل Monotonic. عليه، فإن ارتباط سبيرمان يمكن استخدامه لقياس درجة العلاقة التماثلية بين متغيرين. في أي من الحالتين، فإن ارتباط سبيرمان يمكن بيانه بالرمز r_s لنميزه عن ارتباط بيرسون. إن العملية الكاملة لحساب سبيرمان تحتوي على درجات مرتبة. والمثال التالي يوضح هذه العملية:

إن البيانات التالية تبين العلاقة التماثلية التامة بين X و y. عندما تزيد قيمة X. فإن قيمة y بالتالي تتجه نحو النقصان. ولحساب ارتباط سبيرمان، ينبغي على الباحث بادئ ذي بدء ترتيب القيم المتعلقة بـِ X و y، وبعد ذلك يقوم بحساب ارتباط بيرسون لهذه الرتب.

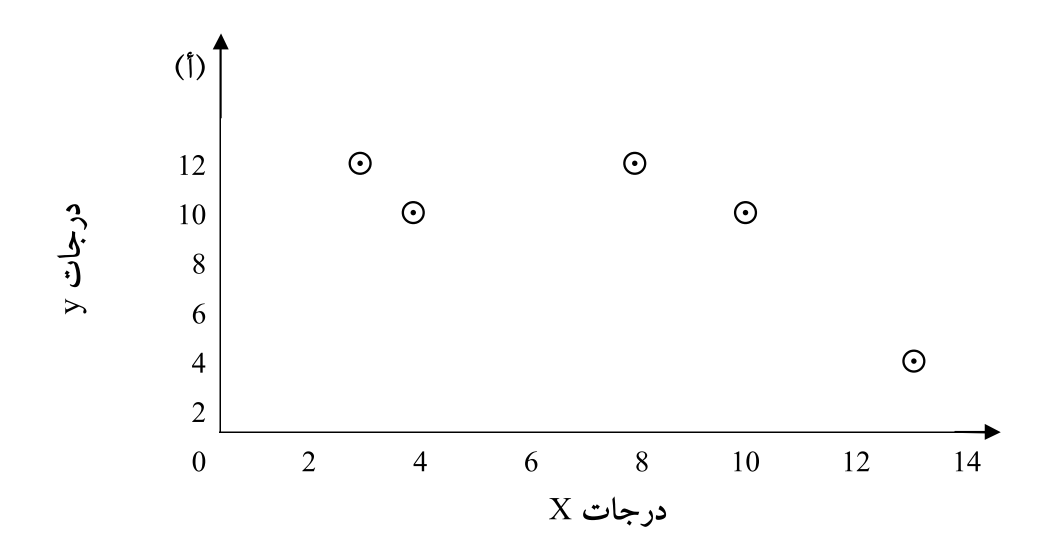
جدول (8-11)

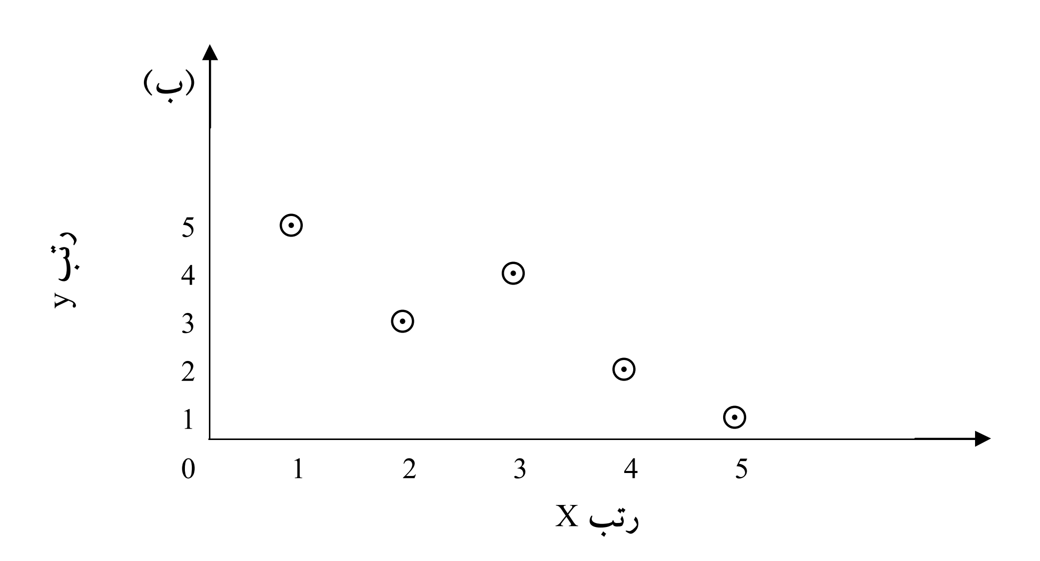
البيانات الأصلية		الرتب		
X	Y	x	Y	xy
3	12	1	5	5
4	10	2	3	6
8	11	3	4	12
10	9	4	2	8
13	3	5	1	5

$$\sum Xy = 36$$

إن شكل الانتشار للبيانات الأصلية والرتب يوضحها الشكل (8 – 5):

ولحساب الارتباط، فإننا نحتاج إلى حساب SS لِـ X و y و Sp. تجدر الإشارة إلى أن كل القيم يتم حسابها مع الرتب، وليس مع البيانات الأصلية. إن رتب X هي ببساطة الأعداد الصحيحة Integers، 1، 2، 3، 4، 5 ويصل مجموع هذه القيم إلى $\sum X = 15$ و $\sum X^2 = 55$.





شكل رقم (8-5) شكل الانتشار والرتب

وأن SS لرتب X هي،

$$SSx = \sum X^2 - \frac{(\Sigma x)^2}{N}$$

$$= 55 - \frac{(15)^2}{5}$$

$$= 55 - 45$$

$$= 10$$

لاحظ أن رتب y مطابقة لرتب X. هذا يعني وجود الأعداد الصحيحة، 1، 2، 3، 4، 5 عليه، فإن SSy ستكون مطابقة لـِ SSx:

$$SSy = \sum y^2 - \frac{(\Sigma y)^2}{N}$$

$$= 55 - \frac{(15)^2}{5}$$

$$= 55 - 45$$

$$= 10$$

ولحساب قيمة SP، فإننا نحتاج إلى: $\sum X$ ، $\sum y$ و $\sum Xy$ للرتب. إن قيم Xy يحتويها الجدول مع الرتب، كما هو واضح في الجدول (11)، إضافة إلى مجموع كل من X_s و y_s الذي يصل إلى 15. وباستخدامنا لهذه القيم نتحصل على:

$$Sp = \sum xy - \frac{(\Sigma x)(\Sigma y)}{N}$$

$$= 36 - \frac{(15)(15)}{5}$$

$$= 36 - 45$$

$$= -9$$

وأن الارتباط النهائي لسبيرمان:

$$r_s = \frac{sp}{\sqrt{(ss_x)(ss_y)}}$$

$$= \frac{09}{\sqrt{(10)(10)}}$$

$$= -0.9$$

ويشير ارتباط سبيرمان إلى أن هذه البيانات تبين اتجاه علاقة قوية سالبة (قريبة من العلاقة التامة).

رتب الدرجات المتعادلة Ranking Tied Scores:

عند تحويل الدرجات الأصلية إلى رتب لارتباط سبيرمان، قد يصادف الباحث درجتان أو أكثر متطابقة. وعندما يكون هناك درجتان أو أكثر متطابقة في القيم، فإن الرتب لهذه القيم ينبغي أن تكون متساوية أيضاً. ولإنجاز ذلك ينبغي إتباع الإجراء التالي:

1- تسجل الدرجات في شكل منظم من أصغر الدرجات إلى أكبرها متضمنة القيم المتعادلة في القائمة.

2- تخصيص الترتيب أول، ثانٍ... الخ، لكل وضع في القائمة المنظمة Ordered List.

3- عندما يصادف أن هناك درجتَيْن أو أكثر تكون متعادلة، هنا ينبغي حساب المتوسط لهذه الأوضاع المرتبة. ويخصص قيمة هذا المتوسط كرتبة نهائية لكل درجة.

إن الإجراء المتبع لإيجاد رتب الدرجات المتعادلة يمكن توضيحه من خلال التالي:

الدرجات	الوضع الترتيبي	الرتبة النهائية	
3	1	1.5	المتوسط لـِ 1، 2
3	2	1.5	
5	3	3	
6	4	5	
6	5	5	المتوسط لـِ 4، 5، 6
6	6	5	
12	7	7	

المعادلة الخاصة بارتباط سبيرمان:

بعد أن يقوم الباحث بترتيب القيم المتعلقة بـِ X والقيم المتعلقة بـِ y، فإن عملية حساب SS، و SP عملية ضرورية جداً، أولاً ينبغي على الباحث أن يلاحظ أن رتب X ورتب y هي حقاً مجرد مجموعة من الأعداد الصحيحة: 1، 2، 3، 4،....... n. ولحساب المتوسط لهذه المجموعة من الأعداد الصحيحة يمكننا الحصول على النقطة الوسطى Midpoint لهذه المجموعة من خلال: M = (n+1) /2. وبشكل مشابه فإنه يمكننا حساب SS لهذه المجموعة من الأرقام الصحيحة من خلال:

$$SS = \frac{n(n^2 - 1)}{12}$$

أيضاً، وبما أن رتب X ورتب y هي متساوية القيم، فإن SSX سوف تكون متطابقة مع Ssy.

ولما كانت حسابات الرتب بالإمكان تبسيطها، ولما كان ارتباط سبيرمان يستخدم بيانات مرتبة، فإن هذه التبسيطات يمكن دمجها في العملية الحسابية النهائية لارتباط سبيرمان. وبدلاً من استخدام معادلة بيرسون بعد ترتيب البيانات، يمكن للباحث أن يضع الرتب مباشرة في المعادلة البسيطة التالية والتي تعطي نفس النتيجة التي تحصلنا عليها باستخدامنا لمعادلة بيرسون:

$$r_S = 1 - \frac{6\Sigma D^2}{n(n^2 - 1)}$$

حيث إنّ:

D الفرق بين ترتيب X وترتيب y لكل فرد. وينبغي الإشارة هنا إلى أن هذه المعادلة الخاصة يمكن استخدامها فقط بعد أن تحول الدرجات الأصلية إلى رتب. وتستخدم هذه المعادلة فقط عندما لا يكون هناك أي رتب متعادلة. أما إذا كان هناك مجموعة قليلة نسبياً من القيم المتعادلة، يمكن في هذه الحالة استخدام هذه المعادلة، الخاصة، لكن هذه المعادلة ستفقد دقتها كلما زادت الأرقام المتعادلة. ولتوضيح تطبيق هذه المعادلة الخاصة نسوق المثال التالي:

الرتب		الفرق	
X	Y	D	D^2
1	5	4	16
2	3	1	1
3	4	1	1
4	2	−2	4
5	1	−4	16

$$\sum D^2 = 38$$

وباستخدام المعادلة الخاصة لارتباط سبيرمان نتحصل على النتيجة التالية:

$$r_S = 1 - \frac{6\Sigma D^2}{n(n^2 - 1)}$$

$$= 1 - \frac{6(38)}{5(25 - 1)}$$

$$= 1 - \frac{228}{120}$$

$$= 1 - 1.90$$

$$= -0.90$$

لاحظ أن هذه النتيجة هي بالضبط نفس النتيجة التي توصلنا إليها من خلال استخدامنا لمعادلة بيرسون بالرتب.

ولتفسير سبيرمان راهو، باعتباره مؤشر لقوة التطابق بين المتغيرات، حيث يقع مداه من 0 (لا يوجد تطابق)، إلى ± (تطابق تام). وأن قيمة التطابق التام تعني عدم وجود أي اختلافات في الرتب بين المتغيرين (إذا كانت كل الحالات قد رتبت تماماً بشكل منظم ومتساوٍ على كلا المتغيرين).

أما قيمة العلاقة السالبة ($r_S = -1.0$) فهي تعني وجود اختلاف تام. (إذا كانت الحالة رتبت بشكل عالٍ على واحد من المتغيرين كانت أقل على المتغير الآخر... الخ).

إنّ قيمة سبيرمان راهو في هذا المثال وصلت إلى $r_S = -0.90$ تشير إلى علاقة قوية بين هذين، وأن إشارة السالب تشير إلى أن هذه العلاقة علاقة سالبة. وإذا ما تم تربيع هذه القيمة $r_S = -0.90^2$ تساوي $r_S = 0.81$ ، وعليه فإن أخطاء التنبؤ التي تم توليدها، تم تخفيض نسبة 81 %، أي أنه عندما نتنبأ برتبة موضوع ما على المتغير التابع (y)، فإن رتب الموضوع الآخر X تم أخذها في الاعتبار.

الخلاصة:

حاولنا في هذا الفصل أن نبين العمليات التي يتم فيها حساب مقاييس التطابق لنسبة التخفيض من الخطأ (PRE) عندما يكون كلا المتغيرين قد تم قياسهما على الأقل على المستوى الترتيبي. وتجدر الإشارة في هذا السياق إلى أنه لا توجد قاعدة سهلة لتحديد أفضلية أي من هذه المقاييس على الآخر.

إن جزءاً من هذه المشكلة يتعلق بمفهوم التطابق نفسه وإن حقيقة مفهوم التطابق يتم تعريفه إجرائياً بطرق مختلفة فعلى سبيل المثال أن مقياس جاما وتاو هي مقاييس متماثلة في حين أن مقياس سومرز d هو مقياس غير متماثل ولذلك فإن الخيار بالدرجة الأولى يتوقف على النموذج الموجه للعلاقة التي يعتقد الباحث فيها. ولذا فإنه من الناحية العملية فإن هذه المقاييس عادة ما تشير إلى نفس الاتجاه وتقود الباحث للحصول على نتائج متشابهة.

أسئلة للمراجعة:

1- إذا كان هناك نقصان في قيمة متغير يرتبط بزيادة في قيمة المتغير الآخر، فما هو اتجاه هذه العلاقة؟

2- لماذا لا يحدث التطابق بين متغيرين بين كونهما إما تطابق موجب أو تطابق سالب عندما يكون على الأقل واحد من المتغيرات تم قياسه على المستوى الاسمي؟

3- من الجداول التالية: احسب عدد الأزواج المتوافقة. مفترضاً أن الأعداد على حاشية كل جدول تشير للقيم على مقياس ترتيبي:

أ-

3	2	1	
12	24	60	1
8	14	32	2

ب-

3	2	1	
12	24	60	1
8	14	32	2

ج-

4	3	2	1	
42	25	17	12	1
24	19	14	10	2
20	16	11	6	3

د-

4	3	2	1	
42	25	17	12	1
24	19	14	10	2
20	16	11	6	3
22	14	9	3	4

4- من الجدول التالي احسب عدد الأزواج غير المتوافقة مفترضاً أن الأعداد في الحاشية تشير إلى قيم مقياس ترتيبي:

أ-

	1	2	3
1	60	24	12
2	32	14	8

ب-

	1	2	3
1	60	24	12
2	32	14	8

ج-

	1	2	3	4
1	12	17	25	42
2	10	14	19	24
3	6	11	16	20

5- من البيانات التالية التي تبين العلاقة بين التدخين ومستوى الصحة لدى مجموعة من الناس:

مستوى الصحة العامة	عادة التدخين		
	لا يدخن	**يدخن**	**المجموع**
سيئة	13	34	47
جيدة بعض الشيء	22	19	41
جيدة	35	9	44
جيدة جداً	27	3	30
المجموع	**97**	**65**	**162**

المطلوب:

أ- بالنظر إلى توزيع الصفوف هل بالإمكان استبيان التطابق بين هذين المتغيرين؟ وما هو اتجاه هذا التطابق؟ هل هذا الاتجاه يظهر عند حساب مقياس التطابق؟

ب-احسب قيمة جاما Gamma، وسومرز Somerś d d وبين ما النتيجة التي توصلت إليها حول العلاقة بين التدخين ومستوى الصحة العامة لدى هؤلاء الناس؟

6- من البيانات التالية احسب قيمة جاما وسومرز d. وفسر هذه النتائج التي تتوصل إليها:

غياب الأم من البيت	مستوى الإنجاز لدى الطفل			المجموع
	ضعيف	جيد	ممتاز	
باقية في البيت (لا تعمل)	20	58	22	100
تعمل جزئياً	15	62	23	100
تعمل طوال النهار	12	62	26	100
المجموع	**47**	**182**	**71**	**300**

7- من البيانات التالية استخدم سبيرمان الخاصة لحساب الارتباط. وأن هذه الدرجات D^2 يساوي 0، 4، 0، 1، 1. وأن مجموع $\sum D^2 = 6$.

الدرجة X	الدرجة y	الرتبة X	الرتبة y	D
5	12	2	2	0
7	18	3	5	2
2	9	1	1	0
15	14	5	4	−1
10	13	4	3	−1

8- من البيانات التالية احسب ارتباط بيرسون مبيناً الخطوات المتبعة؟

الأشخاص	x	Y	Yx	
1	0	0	0	SSX=40
2	2	1	2	SSy=54
3	8	10	80	
4	6	9	54	
5	4	6	24	
	20	30	160	

الهوامش والمصادر:

أولا: الهوامش:

1- George Argyrous, Statistics for Social and Health Research With a Guide to SPSS, Sage Publications ,London, 2001, PP. 173 - 176.

2- Ibid, PP. 179 - 178.

3- Ibid, PP. 180 - 182.

4- Frederick J. Gravetter and Larry B. Wallnau, Statistics for The Behavioral Sciencies, 8th ed, Wadsworth Cengage Learning, USA, 2010, PP. 540 - 547.

ثانيا: المصادر:

1- Frederick J. Gravetter and Larry B. Wallnau, Statistics for The Behavioral Sciencies, 8th ed, Wadsworth Cengage Learning, USA, 2010.

2- George Argyrous, Statistics for Social and Health Research With Guide to SPSS, Sage Publications, London, 2001 .

3- Joseph F. Healey, The Essentials of Statistitics: A Tool for Social Research, Wadsworth Cengate Learning, USA. 2010.

الفصل التاسع

التطابق بين متغيرات تم قياسها على المستوى ذي المسافات والنسبي: شكل الانتشار وخط الانحدار

مقدمة

تناولنا في الفصلين السابقين طرق وصف البيانات عندما يكون أحد المتغيرات على الأقل قد تم قياسه على المستوى الاسمي أو الترتيبي، وقد تم وصف هذه البيانات في شكل جداول تقاطع. ومن خلال جداول التقاطع يتبين لنا التوزيع المشترك لمتغيرين نستطيع من خلالهما ملاحظة ما إذا كان هناك تطابق بين المتغيرين. وعند فحص التوزيعات التكرارية النسبية في الجدول، يقودنا إلى أن نشك في أنَّ هذين المتغيرين يرتبط بعضهما بالبعض الآخر، مما يحدو بنا كخطوة لاحقة بحساب مقاييس التطابق التي تمكننا من الحصول على قيمة رقمية محددة لمثل هذا الشك. وتجدر الإشارة إلى أنه إذا كانت البيانات لهذين المتغيرين تحت الدراسة قد تم جمعها على المستوى ذي المسافات والنسبي، الذي يعني عادة وجود عدد كبير من القيم، الأمر الذي يصبح فيه التعامل مع جداول التوافق غير مناسب لوصف التوزيع. ومن هنا فإن التقنية المرادفة لجداول التوافق للبيانات المناسبة المقاسة على المستوى ذي المسافات / النسبي هي الرسم الانتشاري.

الرسم الانتشاري Scatter Plots:

إنه من الصعوبة بمكان أن نرتب البيانات ذات المسافات المتساوية والنسبية في جداول تقاطع؛ فالبيانات ذات المسافات المتساوية والنسبية عادة لا تقع في عدد صغير من الفئات المنفصلة صغير السن، كبير السن... الخ كالبيانات ذات حجم صغير أو كبير.

إن مثل هذه البيانات بطبيعة الحال يمكن اختزالها إلى قيم قليلة، ولكن هذا الاختزال سيكون على حساب المعلومات. وحيث إن هناك عادةً مجموعة من القيم للمتغيرات التي تم قياسها على المستوى ذي المسافات والنسبي، فإن جدول التوافق سوف يحتوي على عدد كبير من الصفوف والأعمدة، كما توجد عدة قيم في البيانات. وإذا نظرنا إلى التوزيع العمري بالسنوات لسكان بلد ما فإننا نحتاج إلى أكثر من 100 صف من البيانات لكي نأخذها في الاعتبار، فالحقيقة القائلة إن العمر سوف يتوزع على مدى واسع. إن رسم البيانات في شكل انتشاري يمكننا من الحصول على مدى واسع من القيم التي توفر لنا عادة أفضل طريقة لتنظيم مثل هذه البيانات للحصول على انطباع أولي فيما إذا كان هناك وجود لأي علاقة.

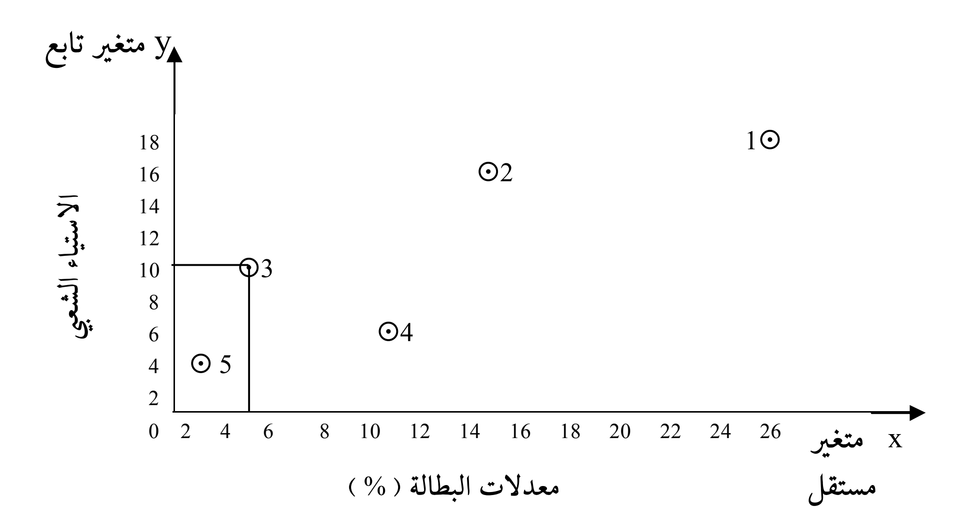
إن رسم الانتشار يبين لنا مجموعة من القيم المتآلفة بحيث إنّ كل حالة تسجل على كلا المتغيرين بشكل تلقائي. إن رسم الانتشار يبين لنا التوزيع المشترك لمتغيرين متصلين ومتناسقين على رسم الانتشار الذي يشير إلى القيم لكل حالة تدون لكل واحد من المتغيرين. فعلى سبيل المثال، قد نرغب في معرفة العلاقة بين معدلات البطالة ومستوى الاستياء الشعبي عن المدن والتي يمكن الحصول عليها من الإحصاءات الرسمية المتعلقة بمعدلات البطالة. فالبطالة لمتغير مستقل X وعدد حالات الاستياء الشعبي لمتغير تابع y لخمس مدن.

جدول (9 ـ 1) معدلات البطالة والاستياء الشعبي في خمس مدن

المدن	معدلات البطالة	معدلات الاستياء
1	25	17
2	13	15
3	5	10
4	10	5
5	2	4

المصدر : George Argyrous,, Statistics for Social and Health Research, With a Guide to Spss, op.cit , p. 233

وبترتيب هذه المعلومات في رسم انتشاري شكل (9 – 1) تجعل من السهولة بمكان قراءتها لتحديد فيما إذا كانت هناك علاقة.



شكل رقم (9 ـ 1) الرسم الانتشاري لبيانات البطالة والاستياء الشعبي

لقد جرت العادة أن يوضع المتغير التابع (y) على المحور الرأسي والمتغير المستقل (X) على المحور الأفقي عند بناء شكل الانتشار. وإذا ما نظرنا إلى أي واحدة من هذه النقاط 1-5 ورسمنا خط مستقيم من الأسفل إلى المحور الأفقي فإنه بإمكاننا أن نجد معدل البطالة في تلك المدينة. وبشكل مشابه برسمنا لخط مستقيم عبر المحور الرأسي يمكننا قراءة عدد حالات الاستياء الشعبي، فالخط المتقاطع لمدينة 3 كما تم توضيحه في هذا الإجراء يشير إلى أن معدلات البطالة لهذه المدينة هو 5 % وأن عدد حالات الاستياء الشعبي هي أيضاً 10 أحداث من الاستياء الشعبي. وبالنظر إلى هذا الشكل فإننا بداهة نرى أن هناك تطابق، لأننا استطعنا أن نتصور ميل الخط ماراً خلال هذه النقاط الخمس. وأن اتجاه التطابق يشار إليه من خلال ما إذا كانت ميول الخط التخيلي أعلى (موجب) أو أسفل (سالب). في هذه الحالة التي أمامنا فإن ميل الخط يكون موجباً مشيراً إلى أن الزيادة في معدلات البطالة متطابقة مع الزيادة في عدد حالات الاستياء الشعبي[1].

خط الانحدار Regression Line:

يعني الانحدار ببساطة مطابقة الخط خلال الرسم الانتشاري للحالات التي تكون مطابقة بشكل جيد للبيانات. وأن أي خط يمكن توضيحه من خلال المعادلة الرياضية التالية المتعلقة بالخط المستقيم. حيث إن: $y = a \pm bX$.

حيث أن: y المتغير التابع.

و X المتغير المستقل.

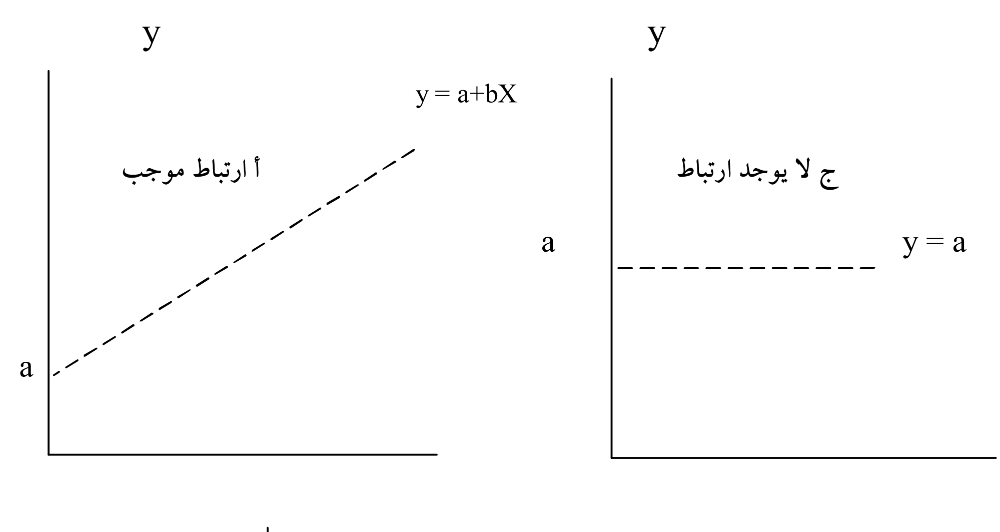
و a قيمة ثابتة وهي المساقة على المحور y من البداية إلى النقطة التي تقطع محور y فهي قيمة y عندما تكون X تساوي 0.

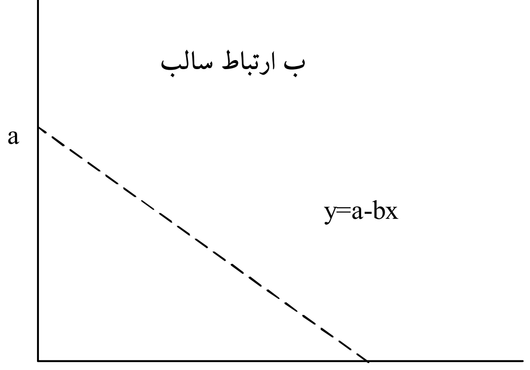
B تشير إلى ميل الخط.

+ تشير إلى التطابق الموجب.

– تشير إلى التطابق السالب.

من خلال هذه المعادلة يمكننا القول بأنَّ الخط المستقيم يمكن تعريفه من خلال عاملين أساسيين أولهما: بداية بنقطة على طول المحور الرأسي، a، وثانيهما هو ميل الخط من نفس هذه النقطة b ± وأن قيمة b هي القيمة التي نرغب فيما إذا كان أي ميل إما أن يكون موجباً أو يكون سالباً، كما تشير لبعض الارتباط بين المتغيرين. ففي الشكل (9 – 2) فإننا نلاحظ ثلاثة خطوط مختلفة تعكس قيمة b في ثلاثة مواقف بديلة لارتباط موجب، وارتباط سالب، ولا يوجد ارتباط.



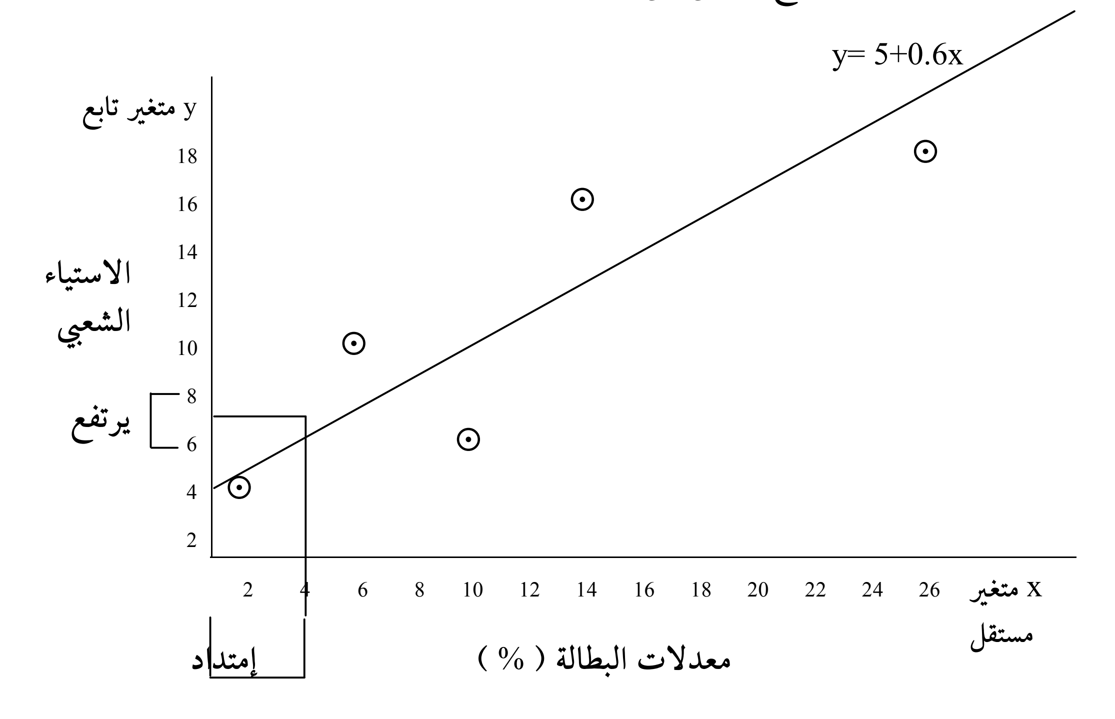


شكل رقم (9 ـ 2): ثلاث خطوات تبين ارتباط موجب أ وارتباط سالب ب ، (ج) لا يوجد ارتباط 0

وبالنظر إلى البيانات المتعلقة بالمدن الخمس يمكننا رسم مجموعة من الخطوط المستقيمة من خلال الشكل الانتشاري، وأن كل واحد من هذه الخطوط سوف يحتوي على معادلته الفريدة والخاصة به. فعلى سبيل المثال، في شكل (9 – 2) قد تم رسم خط يبدو لنا أنه يتوافق مع البيانات بشكل جيد. ويمكن أن نطلق على هذا الخط خط 1 وبدلاً من ذلك يمكننا التعبير عنه بمعادلة رياضية:

$$y = 5 + 0.6X$$

إن السؤال المطروح هو من أين جاءت هذه المعادلة؟



شكل رقم (9-3)

القيمة لِـ: a (5) وهي النقطة على محور y عندما يبدأ الخط. وهي عدد حالات الاستياء الشعبي الذي نتوقع أن نجده في مدينة لديها معدل بطالة صفر (0).

- علامة + تعني أن الخط لديه ميل موجب الذي يشير إلى ارتباط موجب بين هذين المتغيرين.
- أما قيمة 0.6 لـِ b فهي تشير إلى الميل، أو معامل خط الانحدار.

إن معامل خط الانحدار تشير إلى كم من الاستياء الشعبي سيزيد إذا ما زادت معدلات البطالة بـِ 1 % ولما كان الميل لأي خط مستقيم هو الارتفاع على الخط الممتد، فإنه بإمكاننا حساب قيمة (1) لأي ارتفاع في الاستياء الشعبي مثل زيادة 3 بين 5–8. حينئذٍ يمكننا قراءة الزيادة المقابلة في معدلات البطالة، التي تعطينا امتداد خمس (5) وبقسمة الارتفاع على الامتداد (run) فإن الميل سيكون كالتالي:

$$b = \frac{\text{الارتفاع}\,rise}{\text{الامتداد}\,run} = \frac{3}{5}$$

$$= 6.$$

ن الخط الذي أوضحناه الآن يعطينا مدى القيم المتوقعة للاستياء الشعبي اعتماداً على قيمة معدلات البطالة وإن الفرق بين القيمة المتوقعة والقيمة الفعلية للاستياء الشعبي حول معدل بطالة محدد، يطلق عليه المتبقي residual أو مصطلح الخطأ Error Term. إن المتبقي أو مصطلح الخطأ هو الفرق بين القيمة التي نتحصل عليها من خلال المشاهدة للمتغير التابع وبين القيمة المتوقعة للمتغير نفسه بواسطة أو عبر خط الانحدار. لاحظ أن الخط المستقيم لن يمر خلال كل النقاط في الشكل الانتشاري. في حقيقة الأمر أن الخط الجيد من الممكن أنه لا يمس أي من هذه النقاط: عادةً ما تُوجد فجوة بين كل رسم بياني وخط الانحدار باستثناء النقطة التي تقع على الخط ومن هنا تكون لدينا قيمة متبقية. على سبيل المثال، أن الخط الذي لدينا يتكهن بأن المدينة رقم 4 بمعدل بطالة 10 %، وعدد الاستياء الشعبي في هذه المدينة سيكون كالتالي:

$$y = 5 + 0.6X = 5 + 0.6(10)$$

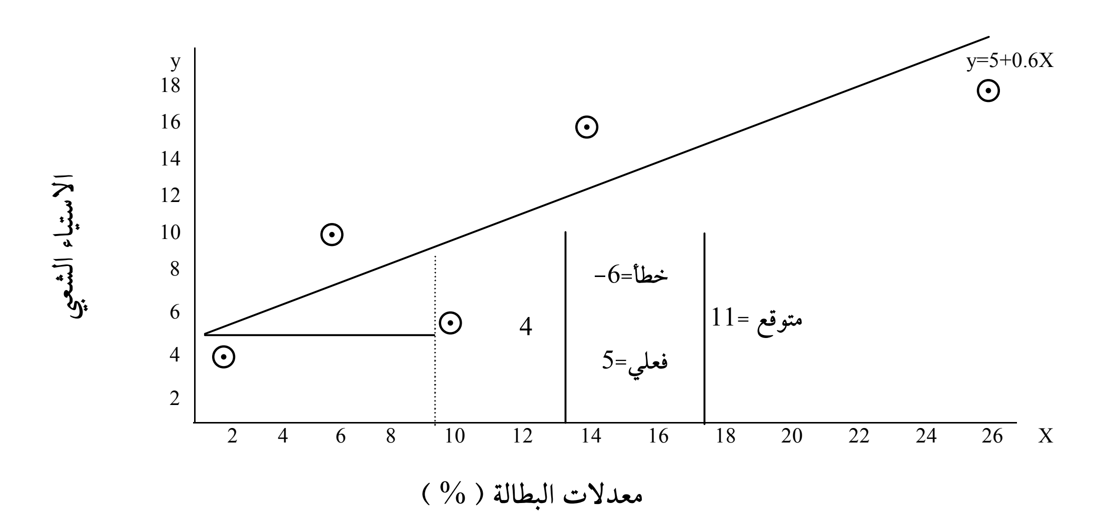
$$= 11$$

وبدلاً من ذلك، يوجد لدينا خمس حالات من الاستياء الشعبي لمدينة 4 بمعدل بطالة يصل إلى 10 %. وبالتالي فإن الخطأ (e) عند هذه النقطة يصل إلى 6-:

المتوقعة الفعلية e = y - y = 5 - 11

= -6

والشكل التالي يوضح هذه المسألة:



شكل (9-4) الدرجات الفعلية والدرجات المتوقعة

واستناداً على الخط الذي تم رسمه فإننا نلاحظ من خلال العين المجردة أن الخط إلى حد بعيد يطابق هذه البيانات. إن أفضل خط هو ذلك الخط الذي يقلل من البواقي ويجعلها قليلة قدر الإمكان. أي الخط الذي يقلل إلى حد أدنى البواقي.

إن تحليل الانحدار يستخدم الفكرة الرئيسية وهي: أقل مربعات الانحدار Least Square Regression (LSR) لتقليص الفارق بين القيمة المتوقعة لـِ y والقيم الفعلية لـِ y (مربعة) على أن يكون هذا الفرق صغيراً قدر الإمكان (عادة ما نربع البواقي أكثر من أن نقوم بجمعها حيث إن مجموع البواقي لأي خط تمر خلاله النقطة التي يكون فيها المتوسط

لكلا المتغيرين التابع والمستقل مساوياً لصفر. ولتقليل أثر العلامة الموجبة والسالبة فإنه يتم تربيع البواقي بحيث يصبح تعاملنا مع الأعداد الموجبة).

إن أقل مربعات الانحدار هي تلك القاعدة التي تبين لنا رسم الخط خلال الرسم الانتشاري والذي يقلل مجموع البقايا المربعة. إنه بإمكاننا إيجاد أقل مربعات خط الانحدار خلال عملية التجربة والخطأ فإنه بإمكاننا أن نستمر في رسم الخطوط خلال شكل الانتشار مع إيجاد المعادلات الخاصة والبواقي حتى نصل في نهاية الأمر إلى الخط الذي يقلل من هذه البواقي. ولحسن الحظ يوجد لدينا بديل عندما نستخدم القاعدتين الاثنتين حتى نستطيع اشتقاق أقل مربعات خط الانحدار بشكل مباشر. إن خط المربعات الصغرى للانحدار ينبغي أن تمر خلال النقطة التي تتساوي فيها المتوسطات للمتغير التابع والمتغير المستقل ($\bar{x}$ و $\bar{y}$). والمتوسط لعدد حالات البطالة $\bar{x}$ تكون:

$$\bar{x} = \frac{\sum x}{N} = \frac{55}{5}$$

$$= 11$$

وأن متوسط معدل الاستياء الشعبي $\bar{y}$

$$\bar{y} = \frac{\sum y}{N} = \frac{55}{5}$$

$$= 10.2$$

وعليه فإن أقل مربعات الانحدار خط الانحدار سوف تمر خلال النقاط المتناسقة (11 و 10.2) ويمكننا تعريف ميل خط الانحدار b من خلال المعادلة التالية:

$$b = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\Sigma(x - \bar{x})^2}$$

وحيث إنَّ هذه المعادلة تبين الفكرة الأساسية التي يحتاجها الخط للتقليل من فروق التربيع بين القيم الفعلية والقيم المتوقعة، من هنا فإنه من السهولة بمكان حساب قيمة b وفقاً للمعادلة التالية:

$$b = \frac{N\,\Sigma(xy) - (\Sigma x)(\Sigma y)}{N\,\Sigma x^2 - (\Sigma x)^2}$$

جدول (9ـ2) حساب ميل خط الانحدار لأقل المربعات

المدن	معدلات البطالة X	حالات الاستياء y	X^2	y^2	Xy
1	25	17	625	289	425
2	13	15	69	225	195
3	5	10	25	100	50
4	10	5	100	25	50
5	2	4	4	16	8
	$\sum X55$	$\sum y = 51$	$\sum x^2 = 923$	$\sum y^2 = 655$	$\sum Xy = 728$

$$b = \frac{N\sum(xy) - \sum(x)\sum(y)}{N\sum x^2 - (\sum x)^2} = \frac{5(728)-(55)(51)}{5(923)(55)^2}$$

$$= \frac{3640-2805}{4615-3025}$$

$$= +\,0.53$$

إن قيمة b يطلق عليها معامل الانحدار، وهي قيمة ذات أهمية لأنها تحدد مقدار أي ارتباط بين متغيرين. فمعامل الانحدار يشير إلى كم من الوحدات في المتغير التابع ستتغير إذا ما تغيرت أي وحدة في المتغير المستقل.

وتجدر الإشارة هنا إلى أنه بعدما حددنا خط الانحدار خلال نقطة محددة (متوسط X و y) وكذلك من خلال إعطاء الاسم الأخير من خلال حساب ميل الخط خلال هذه النقطة، فإنه باستطاعتنا إعطاؤها وصفاً كاملاً من خلال اشتقاق القيمة المتعلقة بـِ a. ومن هنا يمكننا استخدام المعادلة التالية التي تستخدم كلا المَعْلَمَيْن المتعلقين بخط الانحدار الذي

تم بيانه (تلك التي تمر خلال المتوسط المتعلق بـ X والمتوسط المتعلق بـ y وبالتالي يكون لها ميل مساوياً لـ b:

$$a = \overline{y} - b\overline{x}$$

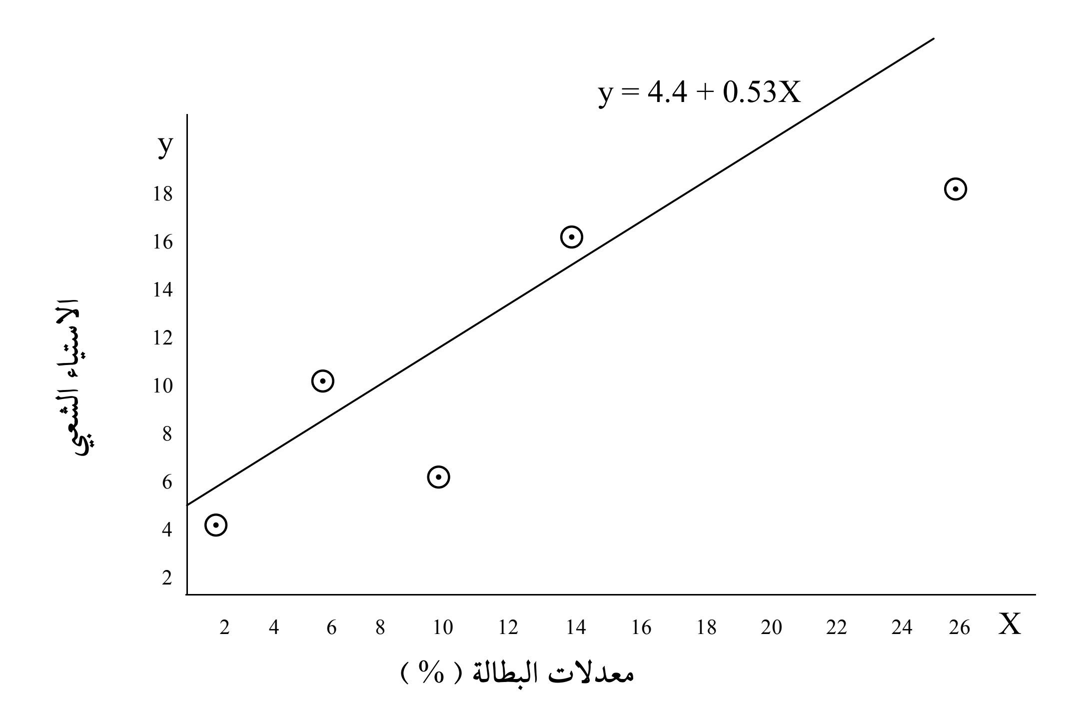
وعليه فقيمة a ستكون:

$$a = \overline{y} - b\overline{x} = 10.2 - 0.53(11)$$

$$= 4.4$$

وعليه يمكننا تعريف أفضل خط مقابل لمثل هذه الحالات من خلال المعادلة التالية:

$$y = 4.4 + 0.53X$$



شكل (9ـ5) خط الانحدار لأقل المربعات (OLS)

السؤال الذي يمكن طرحه هنا هو: ماذا يخبرنا خط الانحدار لأقل المربعات حول العلاقة بين معدلات البطالة وحالات الاستياء الشعبي لمثل هذه المجموعة من الحالات؟ للإجابة على هذا السؤال يمكننا القول:

- توجد علاقة موجبة بين المتغيرين: فالزيادة أو النقصان في معدل البطالة يرتبط مع الزيادة أو النقصان في عدد حالات الاستياء الشعبي.
- يمكننا قياس العلاقة الموجبة: فالزيادة في معدل البطالة لـِ: 1% يرتبط بالزيادة بـِ 0.53 لحالات الاستياء الشعبي.

إنه بإمكاننا الآن استخدام هذه المعادلة لغرض التنبؤ بعدد حالات الاستياء الشعبي في مدينة ما، فمن المرجح أن يكون لديها معدل محدد من البطالة، على سبيل المثال إذا ما علمنا أن مدينة أخرى لديها معدل بطالة يصل إلى 18 % فإن أفضل تكهن يمكن الوصول إليه هو أنها تمر بتجربة 13.9 حالة استياء شعبي [2].

معامل ارتباط بيرسون (r) :

قد تبين لنا من خلال المثال السابق أن قيمة b هي مؤشر ما إذا كانت هناك علاقة بين متغيرين تم قياسهما على مستوى المقياس ذي المسافات والنسبي، وكذلك تبين اتجاه مثل هذا الارتباط. ولكن السؤال هل تشير أيضاً هذه القيمة إلى قوة العلاقة؟ هل قيمة b التي تساوي 0.53 تشير إلى تطابق، قوي أو متوسط أو ضعيف؟ إنَّ هذه القيمة لسوء الحظ لا تشير إلى ذلك.

إن القضية الأساسية هي أن وحدات القياس تتباين من موقف إلى موقف آخر. فعلى سبيل المثال، إذا استخدمنا التناسب بدلاً من نقاط النسب لقياس معدلات البطالة، فعوضاً عن استخدام الحسابات 22، 20، 15، 10، 9 يمكننا استخدام 0.22، 0.20، 0.15، 0.1، 0.09 فإن القيمة المتوقعة لـِ: b ستكون 53 بدلاً من 0.53.

إن العلاقة الفعلية التي نود معرفتها لم تتغير، وإنما تغيرت فقط وحدات القياس. بمعنى آخر، إن قيمة b قد تأثرت ليس فقط بقوة الارتباط، وإنما أيضاً بوحدات القياس. وعليه لا توجد طريقة لمعرفة ما إذا كانت أي قيمة محددة لـِ: b تشير إلى ارتباط ضعيف أو متوسط أو قوي. وللتغلب على هذه المشكلة يمكننا تحويل قيمة b إلى مقياس معياري للارتباط يطلق عليه معامل ارتباط بيرسون (r). إن معامل بيرسون (r) يتراوح مداه دائماً بين −1 و +1 بغض النظر عن الوحدات الفعلية التي تم قياس المتغيرات عليها. والمعادلة المستخدمة لارتباط بيرسون (r) هي:

المعادلة (1)

$$r = \frac{\Sigma(x - \overline{x})(y - \overline{y})}{\sqrt{[(x - \overline{x})^2][(y - \overline{y})^2]}}$$

أو

المعادلة (2)

$$r = \frac{N\,\Sigma(xy) - (\Sigma x)(\Sigma y)}{\sqrt{[N\,\Sigma x^2 - (\Sigma x)^2][N\,\Sigma y^2 - (\Sigma y)^2]}}$$

وإذا نظرنا إلى المثال الذي تم التعامل مع مكوناته من خلال حساب قيمة b جدول (2) فإذا ما استبدلنا الإحصاءات من هذا الجدول وتعويضها بالمعادلة التالية سنتحصل على:

$$r = \frac{N\,\Sigma(yx) - (\Sigma x)(\Sigma y)}{\sqrt{[N\,\Sigma x^2 - (\Sigma x)^2][N\,\Sigma y^2 - (\Sigma y)^2]}}$$

$$= \frac{5(728 - 55(51)}{\sqrt{[5(923) - (55)^2][5(655) - (51)^2]}}$$

$$= \frac{3640 - 2805}{[4615 - 3025][3275 - 2601]}$$

$$= 0.81$$

إما إذا استخدمنا المعادلة الأولى لإيجاد ارتباط بيرسون (r) فإننا نتحصل على نفس القيمة كما هو موضح في التالي:

1	2	3	4	5	6	7
X المدن	y البطالة	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
25	17	+14	+6.8	95.2	196	46.24
13	15	+2	+4.8	9.6	4	23.04
5	10	-6	-0.2	1.2	36	0.04
10	5	-1	-5.2	5.2	1	27.04
2	4	-9	-6.2	55.8	81	38.44
$\Sigma x=55$	$\Sigma y=51$	$\sum(x-\bar{x})=00$	$\sum(y-\bar{y})=00$	167	$\sum(x-\bar{x})^2=318$	$\sum(y-\bar{y})^2=138.8$

$$r=\frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{[(x-\bar{x})^2][(y-\bar{y})^2]}}$$

$$=\frac{167}{\frac{\sqrt{(318)(134.8)}}{\sqrt{42866.4}}}$$

$$=\frac{167}{207.42}$$

$$=0.81$$

إن قيمة r تبين لنا فقط قوة التطابق بل أيضاً اتجاه هذا التطابق. إن قيمة 0.81 تشير إلى الارتباط بين هذين المتغيرين لمثل هذه المجموعة من الحالات فهي قيمة موجبة وقوية.

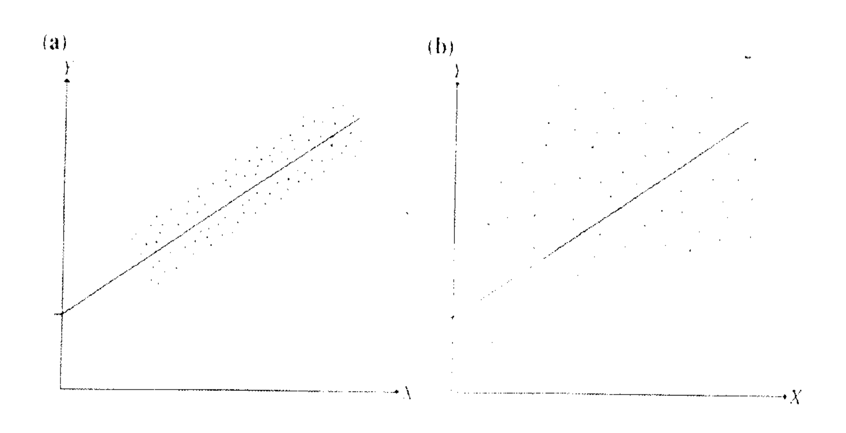
تفسير التباين: معامل التحديد أو التقدير:

قد بينا في هذا الفصل في حينه استخدام خط الانحدار للتكهن بعدد حالات الاستياء الشعبي في مدينة ما من خلال معطى معين لمعدل البطالة. إلا أننا قد رأينا أن هناك دائماً

"هامش للخطأ" في هذا التنبؤ، اعتماداً على مدى قرب النقاط وتجمعها حول خط الانحدار. من هنا يمكننا استخدام خط الانحدار لنبين أن زيادة معينة في X ستقود إلى زيادة كبيرة في y، ولكن إذا كان هناك خطأ كبير بين خط الانحدار والنقاط الفعلية للبيانات فالاحتمال القوي أن التنبؤات سوف تكون خاطئة بقدر كبير أكثر إذا قورنت بالموقف عندما تكون الدرجات متجمعة بشكل واضح حول خط الانحدار.

وكما يظهر في الشكل التالي أنه بالرغم من أن نفس خط الانحدار يتطابق بشكل واضح لكلتا المجموعتين من النقاط فإن ثقتنا ستكون كبيرة في قدرتنا للتنبؤ بقيمة (a) أكثر من قدرتنا على التنبؤ بقيمة (b) ويرجع السبب في ذلك إلى أن خط الانحدار في (a) قادر على تفسير قدر كبير في نسبة التباين في (y) أكثر منه في خط الانحدار (b).

وعليه فإننا بحاجة إلى بعض القياس لكي يبين لنا كم من التباين في المتغير التابع يتم تفسيره بواسطة خط الانحدار.



شكل (9-6)

إنه من حسن الحظ أنه يمكننا أن نقوم بذلك ببساطة من خلال تربيع قيمة (r) للحصول على قيمة معامل التحديد r^2، فالتباين المفسر من خلال خط الانحدار ذا صلة بالتباين المفسر في الحالة التي يوجد بها تطابق.

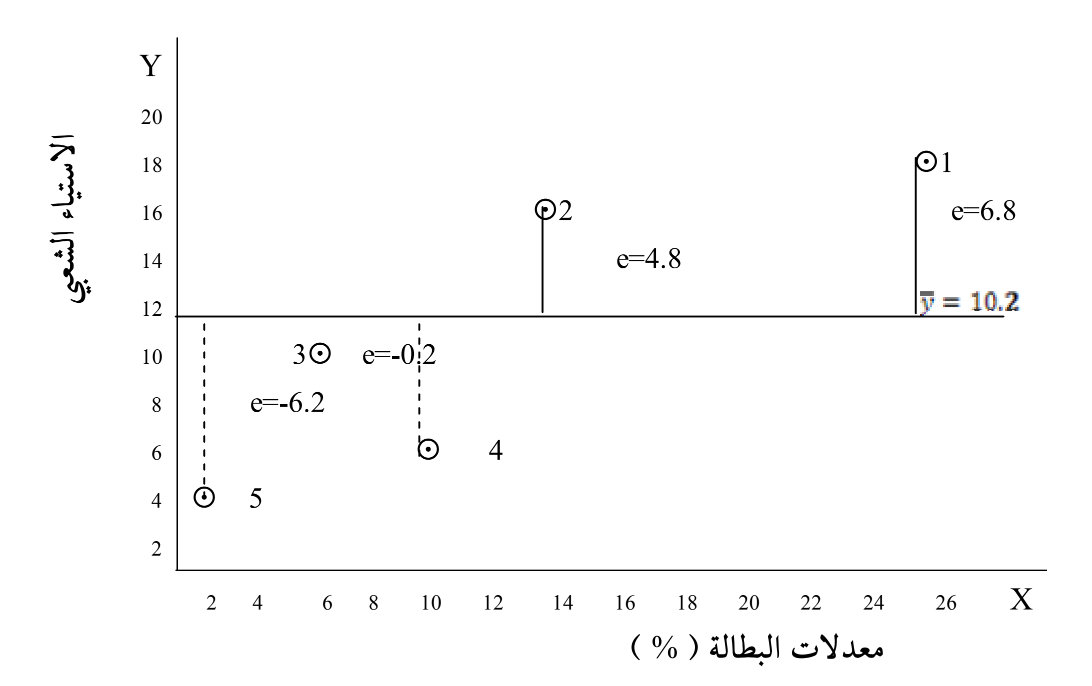
$$r^2 = (0.8!)^2$$
$$= 0.65$$

ويمكننا تفسير معامل التحديد كمقياس غير متماثل للتقليل من نسبة الخطأ في مقياس التطابق أي أن معامل التحديد تشبه إلى حد كبير مقاييس نسبة التقليل من الخطأ Proportional reduction in Error (PRE) التي تم التعرض لها في موضعه في الفصلين السابقين وبالتالي فإن معامل التحديد لديه منطق مشابه لمنطق لامبيدا، إلا أن تطبيقات معامل التحديد تتعلق بالمتغيرات المقاسة على المستويين ذي المسافات والنسبي.

ومن هنا يمكننا القول بأن معامل التحديد يُستخدم لتحديد الدلالة على مقدار مؤاءمة خط أقل التربيعات مع بيانات العينة الذي يسعى الباحث لاختبارها. وتُعرف معامل التحديد بأنه "النسبة بين مجموع مربعات الانحراف الراجح إلى الانحدار إلى مجموع مربعات الانحراف الكلية"[3]، ومن هنا نجد أن معامل التحديد أو التقدير يبين لنا نسبة الاختلاف الكلية في المتغير التابع (y) والتي تم تفسيرها من انحدار y على X.

وتجدر الإشارة إلى أنه أيضاً باستطاعتنا القيام بالتنبؤ حول القيم المتوقعة للمتغير التابع بدون أية معلومات حول المتغير المستقل (X). كما أننا إذاً يمكننا القيام بعملية التنبؤ عندما تكون لدينا معرفة حول المتغير المستقل (X)، ومقارنة معادلات الخطأ فعلى سبيل المثال، إذا أردنا أن نخمن في عدد حالات الاستياء الشعبي في كل مدينة وأن كل ما نعرفه أن متوسط عدد حالات الاستياء الشعبي لكل المدن الخمس هو 10.2، فإن أفضل توقع يمكن الوصول إليه هو أن عدد حالات الاستياء الشعبي لكل مدينة هو 10.2 بغض النظر عن المعدل الفعلي للبطالة [4].

نقوم برسم خط أفقي مستقيم لهذه القيمة كخط انحدار خلال شكل الانتشار. انظر الشكل التالي:



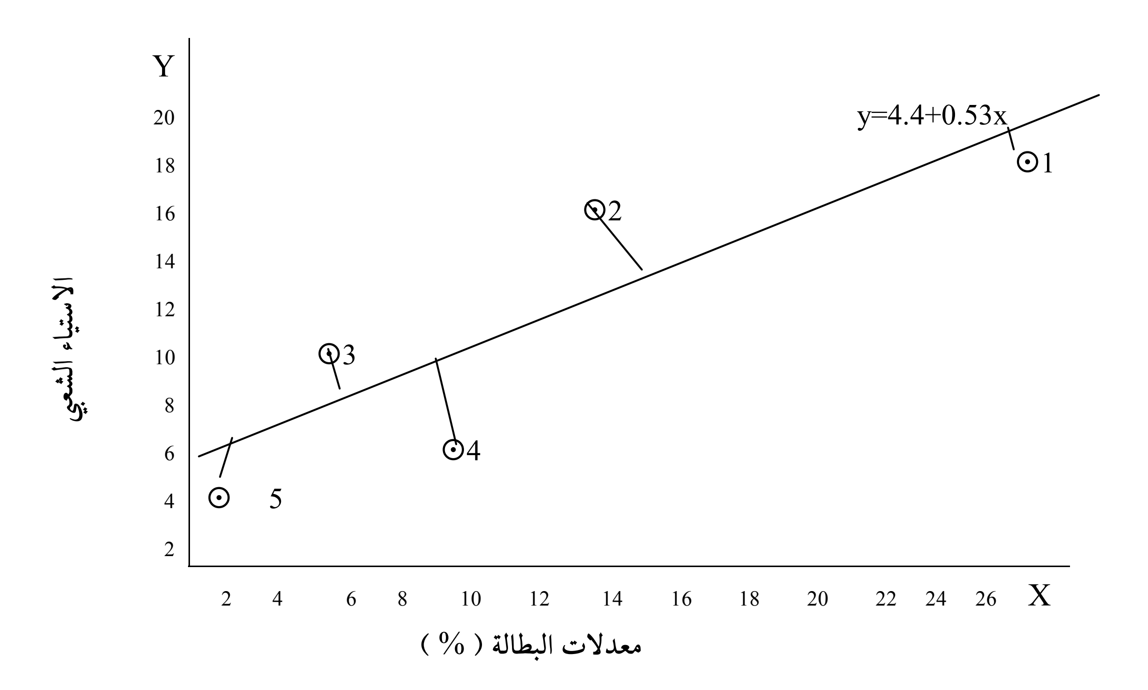
شكل (9ـ7) خط الانحدار بدون معلومات حول المتغير المستقل

إن هذا الخط الأفقي هو الخط الذي تم رسمه عندما لم يكن هناك ارتباط بين هذين المتغيرين (معدل البطالة وعدد حالات الاستياء الشعبي). إن معرفة ما إذا كان معدل البطالة عالياً أو منخفضاً لن يقودنا إلى تغيير توقعاتنا حول حالات الاستياء الشعبي عن المتوسط.

ففي بعض الأحيان قد يأتي هذا الخط بشكل قريب من العلاقة. ففي مدينة 3 يمكننا أن نرى أن هذا الخط يتنبأ بمعدل بطالة يصل إلى 5 %، والتي ستكون 10.2 من حالات الاستياء الشعبي.

فهنا نجد أنه في الحقيقة أن 10 حالات من الاستياء الشعبي. تولد لنا خطأ (e) في هذه المدينة يصل إلى 0.2 فقط. وعلى أية حال، فإنه في شواهد أخرى يمكننا أن نسبب خطأ أكبر باستخدام هذا الخط. ففي مدينة 1 على سبيل المثال، يصل معدل البطالة فيها

إلى 25 % مرة ثانية فإننا نتنبأ بأنه في هذه المدينة تصل حالة الاستياء الشعبي فيها إلى 10.2 لكنه في الواقع نجد أن حالات الاستياء الشعبي في هذه المدينة يصل إلى 17 حالة استياء شعبي. ومن هنا نجد أننا قد ارتكبنا خطأ في التكهن بمعدلات الاستياء الشعبي يصل إلى 6.8. والآن إذا قارنا هذه الأخطاء مع تلك الأخطاء التي فعلناها عندما قمنا بعملية التنبؤ وفقاً لخط انحدار أقل المربعات LSR. فهل هذا الخطأ فعلياً يحسن من توقعاتنا؟ انظر شكل (8).



شكل (9ـ8) خط الانحدار عندما يكون لدينا معرفة بالمتغير المستقل (X)

من هذا الشكل يمكننا أن نلاحظ أنه إذا كان هناك ارتباط وثيق بين هذين المتغيرين، وأن خط انحدار أقل المربعات سيقلل من معدل الخطأ. وبالتالي فإن الهوة بين نقاط هذه البيانات والخط ستكون قليلة جداً باستخدامنا لخط انحدار أقل المربعات أكثر منه باستخدام الخط الأفقي استناداً على فرضية لا يوجد ارتباط.

إن استخدام معامل التحديد كأحد مظاهر خط الانحدار هو الأساس في توضيح ذلك. فقيمة (r^2) تساوي 0.65 تشير إلى أن خط الانحدار لأقل المربعات Least Squares regressim Line تفسر 65% من التباين في المتغير التابع نسبة إلى التباين الذي تم تفسيره من خلال الخط الأفقي وهذه العملية تبين لنا التقليل الجوهري في معدل الخطأ(e).

يتعين علينا هنا أن نميز بين قيمة r و r^2 باعتبارهما قيمتان مرتبطتان بعضهما ببعض. ويعتبر معامل الارتباط مقياس معياري للعلاقة بين متغيرين؛ فمعامل الارتباط يُشير إلى المدى الذي يحدث فيه التغير في أحد المتغيرات يرتبط بالتغير في متغير آخر. عليه فإن قيمة r (تشبه b) التي هي في الأساس أداة للتنبؤ. أما على الجانب الآخر، فإن معامل التحديد يُعتبر مقياساً لتقليل النسبة في الخطأ (PRE)، في كمية التباين المفسر من خلال خط الانحدار، عليه فإن معامل التحديد يمد الباحث بإحساس كم لديه من الثقة يمكن الاعتماد عليها في صحة التنبؤ الذي يقوم به [5].

التباين الكلي: المفسر وغير المفسر:

يمكننا بشكل محدد تعريف التقليل في الخطأ الذي ينتج عندما تأخذ المعلومات حول المتغير المستقل في الاعتبار. هناك مجموعان مختلفان يمكن إيجادهما وبعد ذلك مقارنتهما بالتباين الكلي لِـ y لبناء إحصاء سيشير إلى التحسن في التنبؤ.

1- مجموع التباين المفسر Explained Variance ويوضح لنا مجموع التباين المفسر، يحسن قدرتنا على التنبؤ بـِ y عندما تؤخذ المعلومات المتعلقة بالمتغير المستقل (X) في الاعتبار. ويمكننا الحصول على هذا المجموع من خلال طرح $\bar{y}$ (تكهن درجة y بدون معلومات حول X) من الدرجات المتنبأ بها بواسطة معادلة الانحدار (درجة $\hat{y}$ أو y تم التنبؤ بها عندما كانت المعلومات متوفرة حول X). لكل حالة تربع وتجمع هذه الفروق. ويمكن تلخيص هذه العملية كالتالي: $\sum(\hat{y}-\bar{y})$. وأن النتيجة المتحصل عليها يمكن بعد ذلك مقارنتها بالتباين الكلي في y لنتحقق إلى أي مدى تكون معرفتنا بـِ X قد حسنت من قدرتنا على التنبؤ بـِ y.

والمعادلة المتعلقة بذلك يمكن بيانها رياضياً كالتالي:

$$r^2 = \frac{\Sigma(\hat{y} - \overline{y})^2}{\Sigma(y - \overline{y})^2}$$ = التباين المفسر / التباين الكلي

إن معامل التحديد أو التقدير أو r^2 هي نسبة التباين الكلي في y الذي تم تفسيره من خلال X. وتشير r^2 بالتحديد إلى المدى الذي يساعد متغير X على التنبؤ لفهم أو تفسير متغير y.

2- مجموع التباين غير المفسر: لمعرفة التباين غير المفسر فإننا نقوم بطرح درجات $\hat{y}$ المتنبأ بها من درجات y الفعلية لكل حالة من الحالات؛ وبعد ذلك تربع وتجمع هذه الفروق. ويمكن تلخيص هذه العملية في الآتي:

$$\sum(y - \hat{y})^2$$

إن المجموع المتحصل عليه سيقيس كمية الخطأ في التنبؤ في y التي تبقى حتى بعد أن أخذنا متغير X في الاعتبار. إن نسبة التباين الكلي في y غير المفسرة لـِ X يمكن أن نتحصل عليها من خلال طرح قيمة r^2 من 1.00.

وتجدر الملاحظة هنا أنه كلما كانت العلاقة قوية بين X و y كانت قيمة التباين المفسر أكبر والعكس بالعكس. أنه في حالة العلاقة التامة $r = \pm 1.00$ فإن التباين غير المفسر سيكون صفر وأن r^2 ستكون 1.00 وهذا يشير إلى أن X تفسر كل التباين في y، وأننا بهذا بإمكاننا التنبؤ بقيمة y من X بدون خطأ (e). وعلى الجانب الآخر، فإذا كانت X و y لا يرتبطان بعضهما ببعض ($r = 00.0$)، فإن التباين المفسر سيكون (0) وأن r^2 ستكون كذلك 0.00 في مثل هذه الحالة، فإننا سنتوصل إلى نتيجة مفادها أن X لا يفسر أي شيء من التباين في y وأننا سنكون غير قادرين على التنبؤ بـِ y. ففي المثال الذي تعاملنا معه في هذا الفصل تحصلنا على $r = 81$ وبعد تربيع هذه الفئة تحصلنا على معامل تحديد تصل إلى 0.65. ($r^2 = 0.65$) والتي تشير إلى أن 65 % من التباين الكلي في حالات الاستياء الشعبي في هذه المدن تم تفسيره. ومن هنا نجد أن 35 % من التباين

الكلي في y لم يتم تفسيره من قِبل X، وقد يرجع السبب في ذلك إلى تأثير بعض المتغيرات المؤتلفة، أو إلى خطأ القياس، أو الصدفة العشوائية [6].

القياس والمتغيرات الديميوية Dummy Variobles:

يُعتبر معامل الارتباط والانحدار من أقوى التقنيات الإحصائية، لذلك فإنه في أغلب الأحيان تستخدم هذه التقنيات لتحليل العلاقة بين المتغيرات التي لم يتم قياسها على المستوى ذي المسافات / النسبي. إن هذه الممارسة بشكل عام، لا تمثل أية مشكلة إذا تعلق الأمر بالمتغيرات الترتيبية المتصلة التي لديها مدى واسع من الدرجات المحتملة، بالرغم من أن مثل هذه المتغيرات لا تمتلك نقاط الصفر الحقيقية، ومسافات متساوية من درجة إلى درجة. وقد بينا ذلك عند دراسة ارتباط سبيرمان "راهو" Spearmans` rho. كذلك يستخدم الباحثون في العلم الاجتماعي الارتباط والانحدار عند التعامل مع البيانات الترتيبية المضغوطة Collapsed. إن هذه المتغيرات تمتلك عدداً من الدرجات المحدودة (عادة بين 2 إلى 5)، مثل المسح الذي يبحث في الاتجاه نحو عقوبة الإعدام. إن مثل هذا الانتهاك لمستوى القياس لا يشكل بوضوح مشكلة طالما أن النتائج تعالج بدرجة مناسبة من الحذر.

إن هذه المرونة لدى الباحثين في العلم الاجتماعي ينبغي ألا تتسع لتشمل المتغيرات المقاسة على المستوى الاسمي. إن حساب معامل الارتباط أو الانحدار لمتغيرات مثل الحالة الاجتماعية، أو الطائفة الدينية تعتبر غير ذات جدوى؛ وذلك لأن درجات المتغيرات الاسمية ليست أرقام ولا تمتلك أي خاصية رياضية. إنه بإمكاننا أن نخصص درجة للديانة، الإسلام بدرجة (2) وغير الإسلام بدرجة (1). ولكن الدرجة المتحصلة للديانة الإسلام ليس ضعف الدرجة التي تم تحصيلها للديانة غير الإسلام. وبالتالي فإن الدرجات المتعلقة بالمتغيرات المقاسة على المستوى الاسمي هي مجرد عبارات وليست أرقام.

ولما كانت هذه المتغيرات متغيرات غير رياضية، فإنه من غير المنطق أن نقوم بحساب ميل الخط (b) أو نناقش ما إذا كانت هناك علاقات موجبة أو سالبة.

ومن هنا نجد أن الكثير من المتغيرات المهمة في حياتنا اليومية مثل: النوع، الحالة الاجتماعية، العرق، هي متغيرات اسمية لا يمكن تضمينها في معادلة الانحدار أو التحليل العلائقي، تلك التقنيات الأرقى والأقوى المؤثرة في مجال بحوث العلم الاجتماعي.

تجدر الإشارة إلى أنه من حسن الطالع أن الباحثين قد طوّروا طريقةً لمعالجة هذه المعضلة لكي نضيف المتغيرات المقاسة على المستوى الاسمي وذلك من خلال خلق ما يسمى بالمتغيرات الديميوية Dummy Variables. إن المتغيرات الديميوية يمكن أن تكون على أي مستوى من مستويات القياس بما فيها المتغيرات الاسمية، وإن لديها بالضبط فئتين، الأولى ترمز كـ: (0) والأخرى كـ: (1) وبالتالي ينبغي التعامل بهذه الكيفية مع المتغير المقاس على المستوى الاسمي لكي يكون ذا معني عندما يقدم كمتغير ثنائي يمكن تضمينه لمعادلة الانحدار. على سبيل المثال، يمكننا عند التعامل مع متغير النوع أن يخصص الرمز (0) للذكور، ويخصص الرمز (1) للإناث، العرق: أوروبي (0)، أسيوي (1) (ربما يحتاج الباحث إلى متغيرات ديميوية لإضافة أعراق أخرى أو انتماءات أخرى)؛ الانتماء الديني الإسلام قد يعطى له رمز (2) والمسيحية رمز (0) (مرة ثانية قد يحتاج الباحث إلى متغيرات ديميوية لأخذها في الحسبان لدى ديانات أخرى).

ولتوضيح ذلك، نفترض أننا نقوم بدراسة حول العلاقة بين العرق والتعليم (يقاس التعليم بعدد السنوات التي تم إنجازها). فإذا تم تخصيص رمز (0) للشخص ذو العرق الأبيض، والرمز (1) للشخص ذو العرق الأسود، فإنه بإمكاننا حساب معادلة ميل الخط y . وتكتب هذه المعادلة مستخدمين العرق كمتغير مستقل وحينها نبحث في العلاقة بين هذين المتغيرين. مرة ثانية، نفترض أننا قمنا بقياس عينة مستوى التعليم لدى العرق الأسود وخصصنا رمز (1). للعرق الأسود، والرمز (0) للعرق الأبيض، فإننا سنتحصل على معادلة الانحدار التالية:

$$y = a + bX$$

$$y = (12.0) + (-0.5)X$$

إذا كان التعليم متغير تابع y، والعرق متغير مستقل (X)، فإن خط الانحدار يعبر

العمودي Vertical لشكل الانتشار للنقطة، حيث y=12.0 والقيمة b= −0.5 للميل وهي تشير إلى علاقة سالبة بمعنى أن المبحوث ذو الخلفية السوداء في هذه العينة يسجل متوسط أقل من السنوات الدراسية مقارنة بنظيره الأبيض. وأن النتيجة ستكون واحدة إذا ما تغيرت طريقة الترميز.

فإذا كان الرمز (1) قد أعطي للرجل ذو الأصول العرقية البيضاء، وصفر (0) للرجل ذو الأصول السوداء، فإن قيمة (b) ستبقى بالضبط كما هي، وذلك لأن عملية الترميز المتعلقة بالمتغيرات الديميوية هي عملية اعتباطية، كما هو الحال في المتغيرات الترتيبية. فالباحث يتطلب منه أن يكون واضحاً حول ما تشير إليه القيم المتعلقة بالمتغيرات الديميوية.

كذلك يمكننا أيضاً استخدام معامل بيرسون (r) لتقييم قوة واتجاه العلاقة بالمتغيرات الديميوية. فإذا تحصلنا على سبيل المثال، على قيمة r = −0.23 بين العرق والتعليم، فإننا بالتالي نستطيع القول بأنه توجد علاقة سالبة ضعيفة إلى متوسطة بين هذين المتغيرين لهذه العينة. وباتساق الإشارة المتعلقة بميل الخط (b)، يمكننا أيضاً القول بأن التعليم يقل كلما زادت العرقية أي التحرك من العرق الأبيض إلى العرق السود. كذلك في هذا المقام، يمكننا استخدام معامل التحديد، وذلك بالقول، بأن العرق يفسر 5 % ($r^2 = 0.23^2 = 0.05$) من التباين في التعليم [7].

إجراءات الرسم البياني، والارتباط والانحدار باستخدام SPSS:

أولاً: توليد الرسم الانتشاري Scatter Plot مع خط الانحدار regression Line:

1- من القائمة الموجودة في الجزء العلوي من الشاشة أختر:

Graphs ⇐ Scatter (يعطيك صندوق الحوار Scatter plot).

2- انقر على Define (يعطيك مربع Simple scatter plot).

3- انقر على Number of civil disturbances.

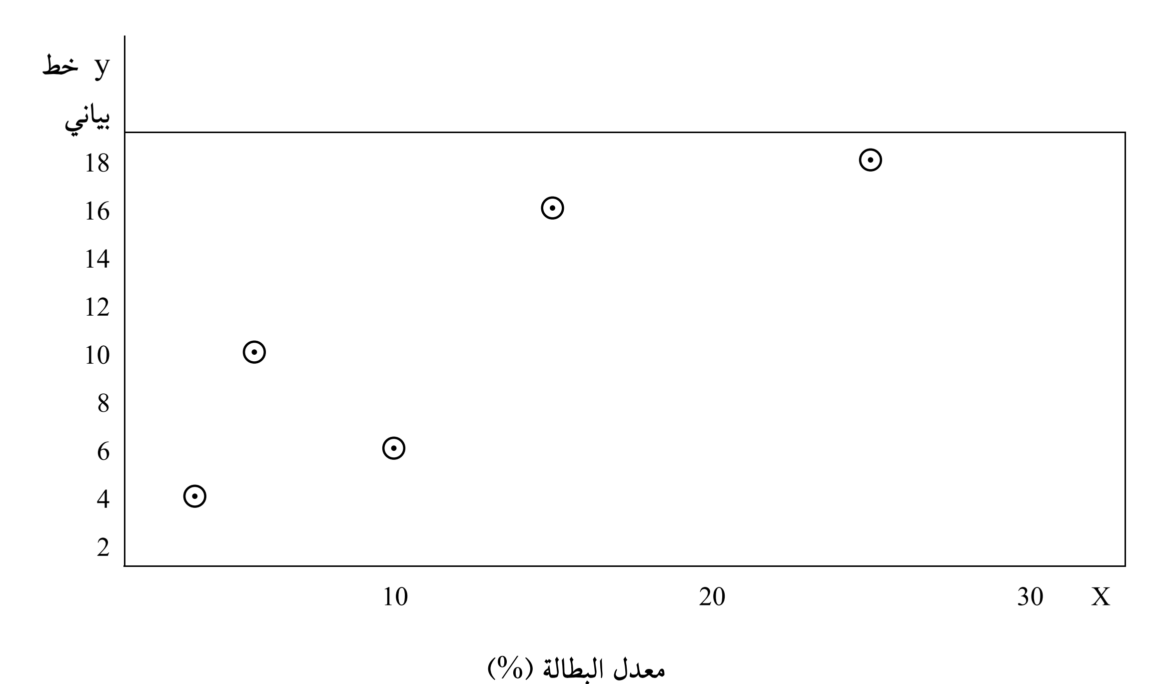
4- انقر على ◄ الذي يشير إلى y Axis في القائمة المحددة للمتغير (هذه العملية ستقود إلى لصق Number of civil disturbances كمتغير تابع يمكن وضعه على y Axis).

5- انقر على unemployment rate (معدل البطالة).

6- انقر على ◄ الذي يشير إلى X Axis في القائمة المحددة للمتغير (هذه العملية ستقود إلى لصق un employment rate كمتغير مستقل يمكن وضعه على X Axis في القائمة المحددة للمتغير المستقل).

7- انقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:



شكل (9ـ9) مخرجات SPSS للرسم الانتشاري

إجراءات إضافة خط الانحدار للرسم البياني باستخدام SPSS:

1- انقر نقرتان على الرسم البياني Scatter plot في نافذة Viewer. هذه العملية تقود إلى الرسم البياني في نافذة Chart Editor Window.

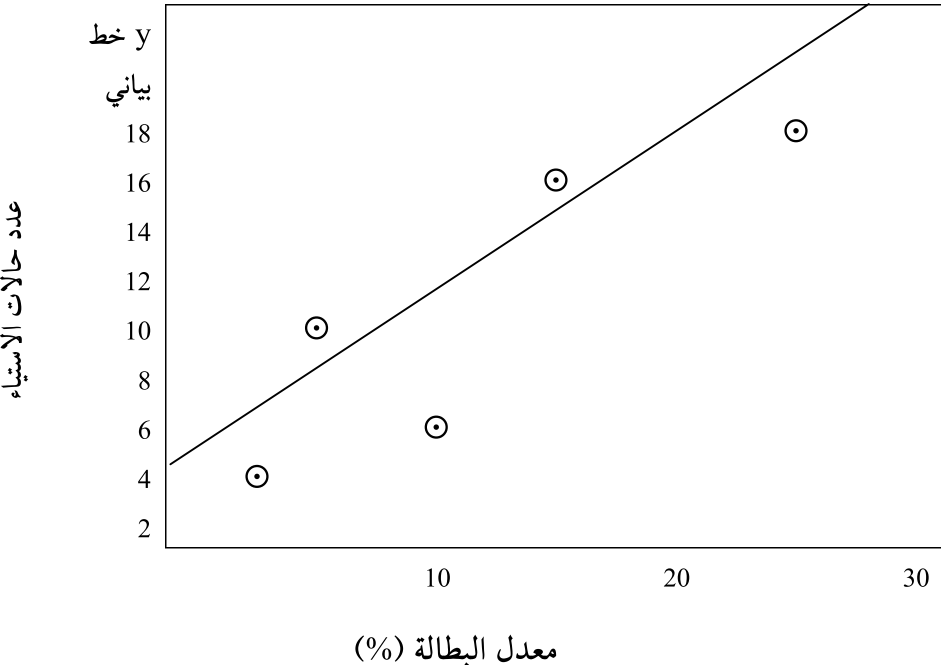
2- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Chart ⟸ Options (هذه العملية تقود إلى Scatter plot options ومن خلال ذلك يمكنك رؤية المساحة المعنونة بـِ: Fit Line التي لديها خيار معنون بـِ Total بعلامة ✓ بجانبه).

3- انقر على علامة ✓ بجانب الـِ Total. هذه العملية تضع علامة ✓ في المربع لبيان أنه تم الاختيار.

4- انقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:



شكل (9 ـ 10) مخرجات SPSS للرسم الانتشار مع خط الانحدار

إجراءات الانحدار Regression مع تقدير المنحنى Curve Estimation باستخدام برنامج SPSS:

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Analyze ⟸ Regression ⟸ Linear

(تقود هذه العملية إلى مربع الحوار Linear Regression dialog box).

2- انقر على Number of civil disturbances.

3- انقر على ◄ التي تشير إلى Dependent في القائمة المحددة للمتغير (هذه العملية ستقود إلى عملية لصق Number of Civil disturbances كمتغير تابع).

4- انقر على Unemployment rate (متغير مستقل).

5- انقر على ◄التي تشير إلى Indepedent(s) في القائمة المحددة للمتغير (هذه العملية ستقود إلى عملية لصق Unemployment rate كمتغير مستقل).

6- انقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

Regression

Variables Entered / Removed[b]

Model	Variable Entered	Variable Removed	Method
1	Unemployment Rate[a]		Enter

a. All requested Variables Entered

b. Dependent Variable: Number of Civil disturbances

Model Summery

Model	R	R Square	Adjusted R Squared	Std. Error of Estimate
1	.807[a]	.651	5.34	3.96

a. Predictors (Constant), unemployment Rate

ANOVA[b]

Model	Sum of squares	Df	MEAN square	F	Sig
Regression	87.701	1	37.701	5.586	.099[a]
Residual	47.099	3	15.700		
Total	134.800	4			

a. Predictors (Constant), unemployment Rate

b. Dependent Variable: Number of Civil disturbances

Correlation[a]

Model	Unstandardized coefficents		standardized coefficents	t	Sig
	B	Std Error	Beta		
(constant)1	4.423	3.013	.807	1.465	.239
Unemployment rate	.525	.222		2.364	.099

a. Dependent Variable: Number of Civil disturbances

المصدر: George Argyrous, op.cit, p. 218

شكل رقم (9 ـ 11) مخرجات SPSS للانحدار

تفسير مخرجات Spss للانحدار:

في مربع ملخص The Model Summery، نجد أن قيمة معامل ارتباط بيرسون (r) تساوي 807. ومعامل التحديد R square تساوي 651. وهاتان القيمتان هما نفس القيمتين اللتين تم حسابهما يدوياً. إن الجزء المهم من مربع معامل الارتباط Coefticients هو العمود المعنون بـِ B تحت معامل اللامعيارية Unstandardized coefficents ويعطينا هذا العمود المعلومات التالية:

- القيمة المتعلقة بـِ y-Intercept (والتي يطلق عليها a في التحليل أعلاه). إلا أن مخرجات SPSS يُطلق عليها كمية ثابتة Constant. وهي 4.423.

- ميل خط الانحدار The Slope of the regression Line الذي يشير إلى معامل الارتباط بمعدل البطالة 525.. مرة ثانية إن هذه النتائج هي نفسها التي تحصلنا عليها عند قيامنا بالعمليات الحسابية اليدوية.

اختبار t لمعامل الارتباط:

في الجزء الأول من هذا الفصل قد تم حساب خط الانحدار، ومعامل الارتباط الإحصائية المرتبطة بمجموعة من الحالات التي تم قياسها على المستوى ذي المسافات والنسبي. وقد تم التعامل مع هذه الإحصاءات الوصفية في إطار استقصاء العلاقة بين معدلات البطالة وحالات الاستياء الشعبي عبر مجموعة من المدن. والنتيجة التي توصلنا إليها كانت كالتالي:

(معدل البطالة) 0.53 + 4.4 = الاستياء الشعبي.

ومن خلال هذه الإحصاءات يمكننا الوصول إلى نتيجة مفادها إلى أن هناك تطابقاً موجباً وقويًّاً بين حالات الاستياء ومعدلات البطالة غير أنه يمكن القول، بأن هذه النتيجة المتحصل عليها من خلال عينة الدراسة قد لا تعكس ما يحدث في كل هذه المدن الخمس.

وكما هو الحال مع أي من الإحصاءات الوصفية الأخرى يمكن حسابها من عينة ما، فإننا بحاجة إلى أن نقرر ما إذا كان معامل الارتباط الذي يصف بيانات هذه العينة يعكس المجتمع الذي سحبت منه هذه العينة. فقد لا يكون هناك ارتباط بين هذين المتغيرين في المجتمع المكون لهذه المدن الخمس (ru)، وإنما يوجد فقط خطأ المعاينة الذي أدى بنا إلى اختيار هذه المدن الخمس التي قد لا تشبه باقي المدن الأخرى. وعليه، فإننا نحتاج إلى إجراء اختبار استدلالي لقيمة معامل الارتباط الذي تحصلنا عليه (r=0.81)[8].

في اختبار t لمعامل ارتباط بيرسون (r) يصاغ الفرض الصفري لهذا الاختبار: إنه لا يوجد ارتباط في المجتمع. في حين يصاغ الفرض البديل، إنه يوجد بعض الارتباط في المجتمع:

$$Ho : ru = 0$$

$$Hi : ru \neq 0$$

وبشكل واضح، فإن معامل ارتباط العينة قد وصل إلى: r = 0.81 وهي قيمة غير متطابقة مع الفرض الصفري. لكن السؤال المطروح هنا هو: هل باستطاعتنا رفض الفرض الصفري الذي مفاده: لا ارتباط في المجتمع على أساس نتيجة هذه العينة؟ ما هي الاحتمالية للحصول على عينة للمدن الخمس بارتباط يصل إلى r = 0.81 بين الاستياء الشعبي ومعدل البطالة من مجتمع يكون الارتباط فيه يساوي صفراً (0). وللحصول على هذه الاحتمالية، ينبغي علينا إجراء اختبار t مستخدمين المعادلة التالية:

العينة $t = \frac{r - ru}{sr}$

$$S_r = \sqrt{\frac{1-r^2}{n-2}}$$

وبالتعويض للقيم المرتبطة بـِ: r و r^2 لهذه المعادلة، فإننا نتحصل على:

$$S_r = \sqrt{\frac{1-r^2}{n-2}} = \sqrt{\frac{1-(0.81)^2}{5-2}}$$

$$= 0.34$$

$$t = \frac{0.81-0}{0.34}$$

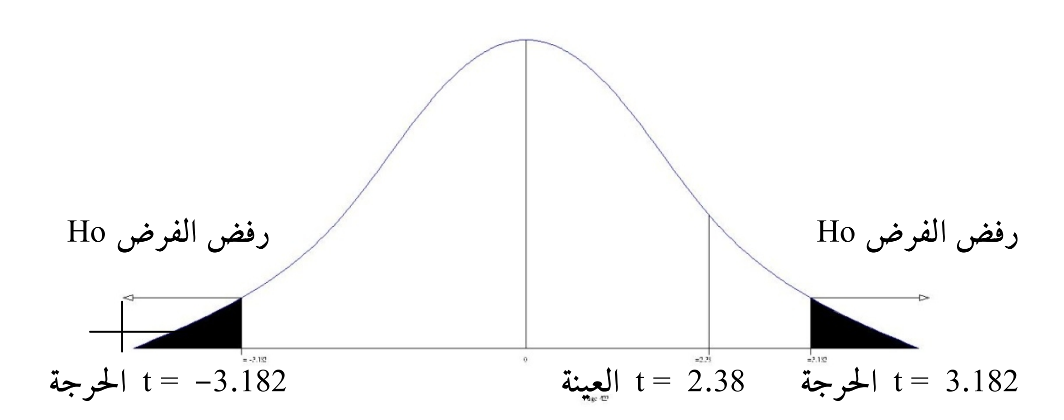
$$= 2.38$$

ولتحديد القيمة الحرجة لـِ t يمكننا الرجوع أو الاسترشاد بالجدول الذي يحتوي على القيم الحرجة لـِ t حيث إنَّ درجة الحرية في هذا المثال تساوي N-2.

	مستوى الدلالة لاختبار أحادي الجانب				
	0.10	0.05	0.02	0.01	0.005
	مستوى الدلالة لاختبار ثنائي الجانب				
	0.20	0.10	0.05	0.02	0.01
df					
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
.	.....	.....	.....	.....	.....
.	.....	.....	.....	.....	.....
.	.....	.....	.....	.....	.....
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.011
14	1.345	1.761	2.145	2.624	2.917
15	1.341	1.753	2.131	2.602	2.947
16	.....	.....	.....	.....	.....
∞	1.282	1.645	1.960	2.326	2.576

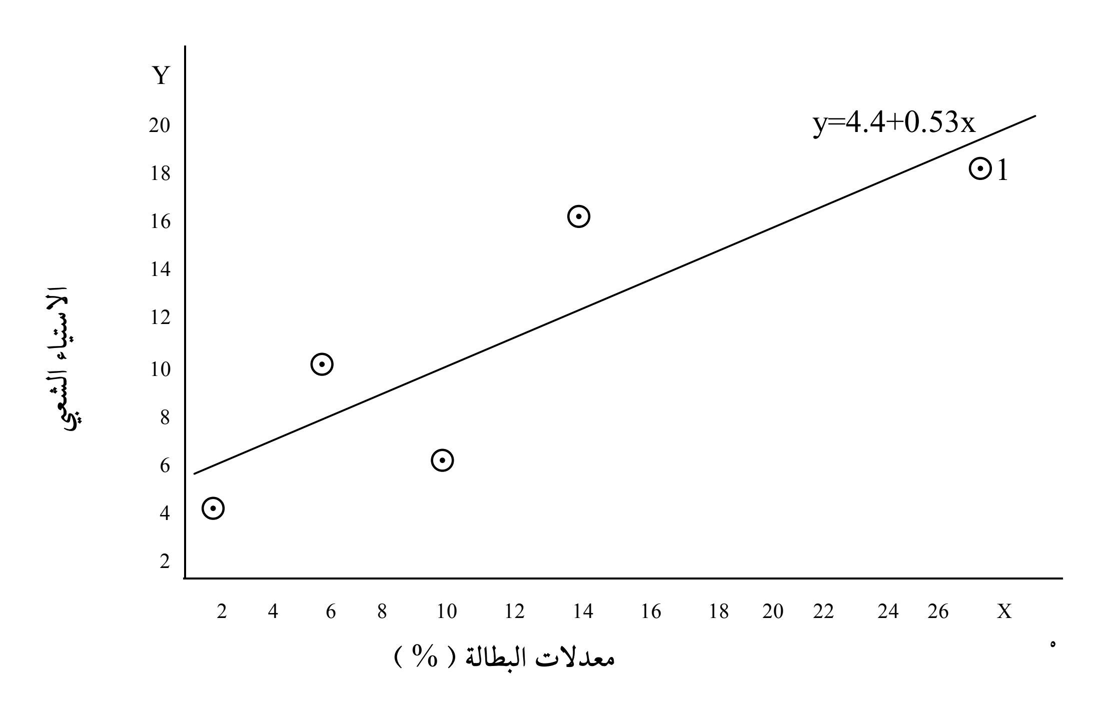
بمستوى ألفا 0.05 واختبار ثنائي الجانب، فإن درجة t الحرجة تكون كالتالي:
$\alpha = 0.05 , df = 3$) t=3.182 الحرجة

وإذا وضعنا هذه المعلومات في شكل بياني (انظر الشكل رقم 11). فإننا بذلك نكون غير قادرين على رفض الفرض الصفري الذي مفاده لا يوجد ارتباط.



شكل (9 ـ 12) الدرجات الحرجة ودرجات العينة

إنه من الأهمية بمكان أن نتوقف لنفكر فيما حدث. ففي العينة التي بين أيدينا تحصلنا على ارتباط قوي بين المتغيرين X و y أي بين معدلات البطالة وحالات الاستياء الشعبي، ولقد أوضح لنا الاختبار الاستدلالي، أنه بالرغم من نتيجة هذه العينة التي ربما قد جاءت نتيجة لعامل الصدفة عند عملية المعاينة من المجتمع حيث إنّ هذين المتغيرين غير مرتبطين. ولنرى لِمَ لَمْ نستطع الإقرار بأن نتيجة العينة تعكس لنا هذه العلاقة في المجتمع؟ فإنه يصبح من الضرورة بمكان أن ننظر إلى شكل الانتشار لهذه البيانات مرة أخرى.



شكل (9ـ13) خط الانحدار لأقل المربعات (LSR)

يمكننا ملاحظة أن خط الانحدار قد تأثر بشكل كبير بالدرجة المتعلقة بمدينة رقم (1) التي يصل معدل البطالة فيها إلى 25 % يقابلها 17 حالة استياء شعبي. ولما كنا نتعامل مع عينة صغيرة (N=5)، فإن الحالة المتطرفة لاشك أنها تؤثر في النتيجة الكلية للعينة. فإذا كانت هذه الدرجة الواحدة مختلفة، كذلك خط الانحدار سيكون هو الآخر مختلفاً. وبما أن هذا الأمر محتمل، فمن غير الممكن أن يكون الارتباط القوي ذا دلالة عند التعامل مع العينات الصغيرة جداً [9].

اختبار الدلالة لارتباط بيرسون (r) باستخدام SPSS:

إجراء اختبار t لمعامل الارتباط الثنائي باستخدام SPSS:

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Analyze ⟸ Correlate ⟸ Bivariate

2- انقر على Number of civil disturbances.

3- انقر على ◄ مشيراً إلى المتغيرات في القائمة المحددة variable(s) هذه العملية تقوم بلصق Number of Civil disturbances في القائمة المحددة للمتغيرات (variable(s).

4- انقر على Unemployment rate.

5- انقر على ◄ مشيراً إلى القائمة المحددة للمتغيرات variable(s) (هذه العملية تقود إلى لصق Unemployment rate في القائمة المحددة للمتغيرات(variable(s)).

6- انقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

Correlations

Correlations

		Unemploment rate	Number of civil disturbances
Unemploment rate	Pearson Correlation sig.(2-tailed) N	1.000 5	.807 5
Number of civil disturbances	Pearson Correlation sig.(2-tailed) N	.807 5	1.000 5

المصدر: George Argyrous, op.cit, p. 425

شكل (9-14) مخرجات الارتباط الثنائي

تفسير اختبار الدلالة لمعامل ارتباط بيرسون (r) من خلال مخرجات SPSS:

إن اختبار الدلالة لارتباط بيرسون (r) يتولد كجزء من المخرجات عندما يجري تحليل الانحدار. إن الإجراء الذي تم إتباعه في الجزء الأول من هذا الفصل لتوليد إحصاءات الانحدار يولد لنا المعلومات الضرورية لإجراء اختبار الدلالة لهذه الإحصاءات.

تجدر الإشارة إلى أن هناك طرق أخرى بديلة من خلالها يمكننا توليد معامل الارتباط بين متغيرين وارتباط درجة t، ومستوى الدلالة.

ومن خلال أوامر الارتباطات بجزمة SPSS الثنائية Bivariate Correlation، يمكن للباحث توليد معامل الارتباط دون المعلومات الإضافية التي تأتي مع تحليل الانحدار الكامل[10].

من خلال النظر إلى مربع مخرجات الارتباط الثنائي يتبين لنا الارتباط بين المتغيرين. أولهما معدل البطالة مع نفسه. وثانيهما عدد حالات الاستياء الشعبي مع نفسه أيضاً. وكل واحد من هذين المتغيرين قد ولدًا معامل ارتباط 1.000. وهذه المعلومات تعتبر ضرورية طالما أن أي متغير بالضرورة يرتبط مع نفسه.

ففي الصف الأول من المربع يوضح لنا الارتباط بين معدل البطالة وحالات الاستياء الشعبي (807.)، وأن دلالة هذا الارتباط تصل إلى 099. وهذا يشير إلى أنه بالرغم من أن معامل الارتباط عالٍ، فهو ليس دالاً على مستوى (05.). وعليه يمكننا القول بأن هذه النتيجة جاءت كنتيجة لتباين المعاينة. أما الصف الثاني من المربع فيعطينا معامل الارتباط بين عدد حالات الاستياء الشعبي ومعدل البطالة وهو نفس المعامل المبين في الصف الأول من المربع.

اختبار t لمعامل ارتباط سبيرمان rho:

من خلال المثال السابق تعرفنا على الإجراءات المتعلقة بحساب معامل ارتباط بيرسون (r) ببحث ارتباط بيرسون (r) في استقصاء التطابق بين متغيرين تم قياسهما على المستوى ذي المسافات والنسبي. أما فيما يتعلق بارتباط سبيرمان، فإن هذا الإجراء يكون

ملائماً عندما يكون أحد المتغيرات على الأقل تم قياسه على المستوى الترتيبي. وإذا ما نظرنا بشكل دقيق إلى الإجراءات المتبعة في حساب هاذين النمطين من معامل الارتباط، فإننا نجد أنهما تقريباً متماثلان في الإجراءات، إلا أن الفرق بينهما هو أن معامل ارتباط (r) يمكن حسابه من البيانات الخام، في حين يعتمد ارتباط الرتب في حسابه على البيانات المرتبة.

تجدر الإشارة إلى أن اختبار الدلالة لكلا النمطين هو اختبار واحد، بمعنى، أن المعادلة لحساب درجة t للعينة هي واحدة بغض النظر، عما إذا كنا نختبر دلالة معامل ارتباط بيرسون (r) أو معامل الرتب $r_S(rhO)$.

$$\text{العينة} \quad t = \frac{r_S - P}{S_r}$$

$$S_r = \sqrt{\frac{1 - r_S^2}{n - 2}}$$

إذا أراد الباحث حساب قيمة $r_S(rhO)$ فعليه أولاً حساب الفرق في الرتب لكل حالة من الحالات المدروسة، (D)، وبعد ذلك تربيع هذه الفروق. وبعد إنهاء هذه العمليات الحسابية ينبغي عليه تعويض تلك العمليات الحسابية في معادلة سبيرمان راهو (rhO).

$$r_S = 1 - \frac{6\Sigma D^2}{n(n^2 - 1)}$$

فإذا توصلنا إلى ارتباط بين متغيرين يصل إلى -0.8 = r_S من بيانات تحتوي على $\sum D^2$ = 1225.5 لعدد 16 حالة.

فإننا بذلك يمكننا اتباع الإجراءات الخمسة المتعلقة باختبار الفروض:

الخطوة الأولى: صياغة الفرض الصفري والفرض البديل:

Ho: لا يوجد ارتباط بين X و y

$$Ho: P = 0$$

Hi: يوجد ارتباط بين X و y

$$Ho: P \neq 0$$

الخطوة الثانية: اختيار اختبار الدلالة:

حيث إننا نبحث في الارتباط بين متغيرين تم قياسهما على المستوى الترتيبي. وإن البيانات تم وصفها من خلال حساب ارتباط سبيرمان rho للرتب. فإن الاستدلال المناسب هو اختبار t لمعامل الارتباط.

الخطوة الثالثة: حساب درجة العينة:

لقد تم حساب معامل ارتباط سبيرمان راهو بين المتغيرين X و y لعدد 16 حالة، وجاءت قيمة الارتباط $r_S = -0.80$

$$r_S = \frac{6\sum D^2}{n(n_2 - 2)} = \frac{1225.5}{16(16^2 - 2)}$$

$$= -0.80$$

ولكي نرى ما إذا كانت هذه النتيجة قد جاءت للتباين العشوائي random variation عند إجراء عملية المعاينة من مجتمع حيث إنّ هذين المتغيرين لا يرتبطان بعضهما ببعض، فإننا بالتالي نحتاج أولاً إلى حساب الخطأ المعياري لتوزيع المعاينة لِ rho.

$$S_r = \sqrt{\frac{1 - r_S^2}{n - 2} = 1 - \frac{(-0.80)^2}{16 - 2}} \; .$$

$$= 0.16$$

وعليه، فإن درجة t للعينة ستكون كالتالي:

$$t = \frac{r_S - P}{S_r} = \frac{-0.80}{0.16}$$ العينة

$$= -5$$

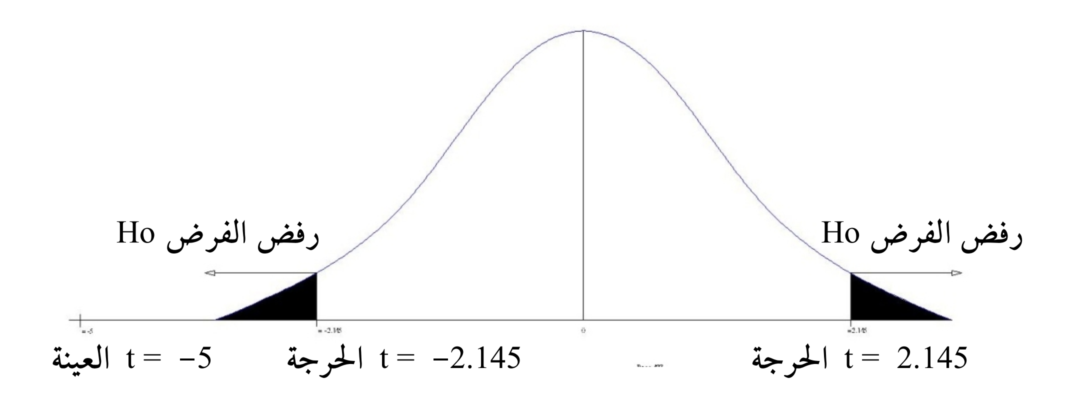
الخطوة الرابعة: حساب الدرجة (الدرجات) والمنطقة الحرجة:

بدلالة 0.05 = ∝ لاختبار ثنائي الجانب، يمكننا الإشارة إلى جدول توزيع t للقيم الحرجة لكي تقرر القيمة الحرجة لـِ t بدرجة حرجة df = 16 – 2 = 14 .

$$t = \pm 2.145 \ (\propto = 0.05 \, , df = 14)$$

الخطوة الخامسة: اتخاذ القرار:

إذا ما قمنا برسم درجات العينة والدرجات الحرجة في الشكل البياني التالي، فإنه يمكننا رفض الفرض الصفري. وعلى أساس هذه العلاقة القوية التي تـحصلنا عليها من خلال العينة المدروسة، فإنه ليس بإمكاننا القول بأنه لا يوجد ارتباط في المجتمع [11].



شكل (9ـ15) درجة العينة والدرجة الحرجة

أسئلة للمراجعة:

1- لماذا ينبغي على الباحث رسم شكل الانتشار للبيانات قبل إجراء تحليل الانحدار؟

2- ماذا تشير a و b في تحليل الانحدار؟

3- من خلال معادلات الانحدار بين اتجاه العلاقة؟

$$y = 30 + 42 X$$

$$y = 30 - 0.38 X$$

$$y = -0.5 + 0.38 X$$

$$y = -0.5$$

$$y = -0.5 X$$

4- اشرح الفرق بين ارتباط بيرسون (r) ومعامل خط الانحدار (b).

5- من البيانات التالية ارسم شكل الانتشار؟

a)

X	5	6	9	10	10	13	15	18	22	27
y	35	26	30	22	28	28	20	21	15	18

b) من خلال هذا الرسم الانتشاري، بين الإشارة التي تسبق المعامل (على سبيل المثال، علاقة موجبة أو علاقة سالبة).

6- ما هي توقعات أقل المربعات Least Sqnare لِـ y عندما تكون X = 12 = y = 27.165 - 0.15X.

7- إذا كانت معادلة خط الانحدار y = 40 + 0.7X لخط انحدار تم رسمه لبيانات متعلقة بعدد سنوات الحياة المتوقعة Life expectancy ومصروفات الدولة على الرعاية الصحية لكل فرد من السكان لمجموعة من الدول النامية. وقد اعتمد عدد سنوات الحياة المتوقعة كمتغير تابع. ومصروفات الدولة على الرعاية الصحية كمتغير مستقل. وأن معادلة خط الانحدار كالتالي: y = 40 + 0.7X.

- ما هو عدد سنوات الحياة المتوقعة إذا قامت الدولة بصرف مبلغ قدره 30.000 دينار على الرعاية الصحية لكل فرد؟
- ما هو عدد سنوات الحياة المتوقع إذا لم تقم الدولة بصرف أية مبالغ مالية على الرعاية الصحية عندما تكون y = 40، X = 0.
- هل يمكنك القول بأنه توجد علاقة قوية بين المتغيرين؟

8- البيانات التالية تبين عدد الأطفال ومساهمة الزوج في أعمال المنزل (بيانات تصورية):

العائلة:	1	2	3	4	5	6	7	8	9	10	11	12
عدد الأطفال:	1	1	1	1	2	2	3	3	4	4	5	5
عدد الساعات الأسبوعية التي يقضيها الزوج في العمل المنزلي	1	2	3	5	3	1	5	0	6	3	7	4

- أوجد شكل الانتشار وماذا يعني ذلك؟
- أوجد قيمة a، b.
- ارسم معادلة خط الانحدار y على X.

9- مسح اجتماعي توصل إلى معامل ارتباط بين عدد السنوات الدراسية التي يقضيها المبحوث، واتجاهاته نحو تولي المرأة مهام إدارية عليا.

قيمة الارتباط (r) تساوي r = 0.54، وحجم العينة المدروسة في هذا المسح كانت 140 حالة، ودلالة هذا الارتباط كان اختبار t الذي وصلت قيمته إلى 7.54.

ما هي النتيجة التي يمكنك الوصول إليها حول طبيعة هذه العلاقة بين هذين المتغيرين؟

10- من البيانات التالية: ارسم شكل الانتشار ؟ أوجد معادلة خط الانحدار y على X.
إذا كانت X = 16. ما هي القيمة التقديرية لـِ y.

$$b = 0.87, a = 3.775$$

$$\hat{y} = \text{---} + \text{---}\ X$$

X	18	14	10	15	7	12	13	8	9	17	15	12
y	20	11	14	16	10	10	17	11	12	20	18	12

الهوامش والمصادر:

أولا: الهوامش:

1- George Argyrous, Statistics for Social and Health Research, With a Guide to Spss Sage Publications ,London, 2001, P. 201.

2- Ibid, PP. 201 - 209.

3- سعد اللافي، الإحصاء الاستنتاجي، ج 1، منشورات أكاديمية الدراسات العليا، طرابلس، 2003، ص 379.

4- George Argyrous, op.cit., PP. 209 - 212.

5- Ibid, P. 213.

6- Joseph F. Healey, The Essentials of Statistics: A TOOL for Social Research, 2[ed], Wadsworth Cengage Learning, UAS, 2010, PP. 343 - 345.

7- Ibid, PP. 349 - 350 and George DiEkhoff, Statictics for the social and Behavioral Sciences: Univariate, Bivariate, Multivariate, the Megraw Hill, Companies, Inc. USA, 1992, P.279.

8- George Argyrous, op.cit., P. 421.

9- Ibid, P. 424.

10- Ibid, P. 424.

11- Ibid, PP. 425 - 428.

ثانيا: المصادر:

1- Frederick J. Gravetter and Larry B. Wallnau, Statistics for The Behavioral Sciences, 8th ed, USA, 2010.

2- George Argyrous, Statistics for Social and Health Research, With a Guide to Spss, Sage Publications ,London, 2001.

3- George DiEkhoff, Statictics for the social and Behavioral Sciences: Univariate, Bivariate, Multivariate, the MCGraw- Hill, Companies, Inc. USA, 1992.

4- Joseph F. Healey, The Essentials of Statistics: A TOOL for Social Research, Wadsworth Cengage Learning, UAS, 2010.

5- سعد اللافي، الإحصاء الاستنتاجي، ج 1، منشورات أكاديمية الدراسات العليا، طرابلس، 2003.

الجزء الثالث

الإحصاءات الاستدلالية البارامترية

[حالة العينة الواحدة]

- الفصل العاشر: توزيعات المعاينة ـ التقدير، وفترات الثقة
- الفصل الحادي عشر: اختبار الفروض: اختبار Z لمتوسط عينة واحدة
- الفصل الثاني عشر: اختبار t لمتوسط حسابي لعينة واحدة

الفصل العاشر

توزيعات المعاينة - التقدير، وفترات الثقة

أولاً: توزيعات المعاينة:

تلعب الإحصاءات الوصفية دوراً كبيراً في تلخيص التوزيعات عموماً تلك التوزيعات التي تساعد في الإجابة على السؤال المطروح. فإذا كانت مجموعة الحالات التي تم أخذها كمقياس تشمل كل الحالات المحتملة التي نرغب في قياسها "المجتمع الإحصائي" فإن عملية البحث سوف تنتهي بالعملية الحسابية لهذه المقاييس الوصفية فالاستقصاء الذي يشمل كل عضو من السكان يُطلق عليه الإحصاء السكاني وأن الإحصاءات الوصفية لهذا المجتمع يُطلق عليها معلمات. فالمعلمة: هي ملخص وصفي لمتغير محدد في المجتمع فمتوسط الدخل لكل العائلات في مدينة بنغازي مثلاً هو معلمة. كذلك الأمر في التوزيع العمري لسكان مدينة بنغازي. وعندما يقوم الباحث بالتعميم من العينة فهو يستخدم ما يلاحظه من خلال العينة للتكهن بمعلمات المجتمع. وعند استخدام المعادلة الرياضية فإن المعلمات تشير إلى u كمتوسط للمجتمع و σ كانحراف معياري.

وتجدر الإشارة إلى أنه في بعض الأحيان تكون لدينا معلومات حول المجتمع ككل كالإحصاءات التي تجريها دولة ما ومن خلال هذه الإحصاءات يمكننا معرفة التوزيع

العمري لكل السكان في فترة تاريخية محددة. إلا أننا في وقت آخر قد لا يكون في إمكاننا معرفة التوزيع العمري لمجتمع ما، بالرغم من وجود هذا التوزيع. وعليه، فإننا في تلك الحالة نقوم بإجراء بحوث في الغالب نتعامل مع جزء من المجتمع أي عينة. وتسمى المقاييس الوصفية التي تستخدم لتلخيص المعلومات المتعلقة بالعينة بإحصاء العينة: وهو الملخص الوصفي لمتغير في عينة تستخدم للتكهن بمعلمة المجتمع. ويشار إلى إحصاء العينة بالتعبير الرياضي ب: $\overline{X}$ الذي يشير إلى متوسط العينة و S للانحراف المعياري لهذه العينة.

توجد لدى الباحث عدة أسباب تدعو إلى سحب عينة من المجتمع بدلاً من إجراء المسح الشامل لكل مفردات المجتمع وهذه الأسباب:

- قد يكون من الصعوبة بمكان على الباحث أن يحصر كل أفراد المجتمع إما لعدم توفر قائمة تحتوي على كل مفردات المجتمع أو أن بعض أفراد المجتمع لا يرغبون في المشاركة في مثل هذه الدراسة.

- قد تكون المعاينة في بعض الأحيان أكثر دقة فإذا كان لدينا أي سبب أن نعتقد بأن عملية المسح عن طريق العينة تولد أخطاء عندئذٍ فإن إجراء الإحصاء الشامل يمكن أن يوسع من هذه الأخطاء. فعلى سبيل المثال، فإن الإحصاء الشامل يتطلب فريقاً بحثياً كبيراً لإجراء المسح يمكن أن يقود إلى جامعي بيانات ليست لديهم الخبرة في جمع البيانات، في حين أن الفريق الصغير يمكن أن يكون مدرباً تدريباً جيداً ولديه خبرة كبيرة في جمع البيانات، ومهما كانت هذه الأسباب، فإن المشكلة الأساسية التي تظهر هي: هل الإحصاءات الوصفية التي نتحصل عليها من خلال العينة مساوية للإحصاءات التي نتحصل عليها إذا ما قمنا بإحصاء دقيق وشامل. هل إحصاء العينة يكون ممثلاً للمجتمع الذي سحبت منه هذه العينة؟

لاشك أن الباحث تكون لديه القدرة على سحب عينة ممثلة من المجتمع وأن عملية التباين العشوائي قد تؤثر في العينة من حيث التمثيل. والسؤال المطروح في هذا السياق هو: ما هي الأسس التي على ضوئها يمكننا أن نقوم بتعميم صادق من العينة على المجتمع؟ فعلى سبيل المثال، يمكننا أن نسحب عينة من 120 فرداً من منطقة جغرافية محددة

ونطرح على كل واحد منهم سؤالاً مفاده: كم يبلغ من العمر بالسنوات؟ ففي هذه الحالة يكون متغير العمر هو الأساس وقد تم قياسه على المستوى ذي المسافات والنسبي وأنه بإمكاننا أن نصف المعلومات التي تحتويها هذه البيانات من خلال حساب مقياس النـزعة المركزية لكي نتحصل على توزيع الدرجات، وبحساب مقياس التشتت يكون لدينا الإحساس بالدرجات حول المتوسط. ومن خلال الرسم البياني سيعطينا عموماً الانطباع حول التوزيع، إنه يمكننا الشعور بأن هذه الطرق ليست الوحيدة لوصف التوزيع كما بيّنا. لكن في الغالب قد نفي بمتطلبات الكثير من الأسئلة البحثية. إن هذه المعلومات قد تكون ذات أهمية في ذاتها إلا أننا قد نقوم بجمع بيانات حول عينة ما لأن لدينا سؤالاً آخرَ نود طرحه: ما هو المتوسط العمري لكل الناس في هذه المنطقة فإذا كان المتوسط العمري لهذه العينة هو 36 سنة فالسؤال الذي يمكن طرحه هو: هل بإمكاننا التعميم من هذه العينة على المجتمع ككل. هنا تظهر عملية التباين العشوائي التي قد يتسبب في إعاقة قدرتنا على التعميم الصادق من خلال إحصاء العينة.

كيف لنا أن نكون واثقين بأن العينة التي قمنا بسحبها لم يتم سحبها عن طريق الصدفة لتحتوي على نسبة قليلة من كبار السن أو نسبة قليلة من الصغار فيما يتعلق بالمجتمع. قد نطرح هذه المشكلة من خلال استخدامنا لإحصاء الاستدلال، فإحصاءات الاستدلال هي تقنيات عددية تهدف للوصول إلى نتائج حول المجتمع استناداً على البيانات التي تم الحصول عليها من عينة عشوائية تم سحبها من ذلك المجتمع. ولإجراء الإحصاء الاستدلالي فإنه يمكننا توليد ثلاث مجموعات عددية منفصلة:

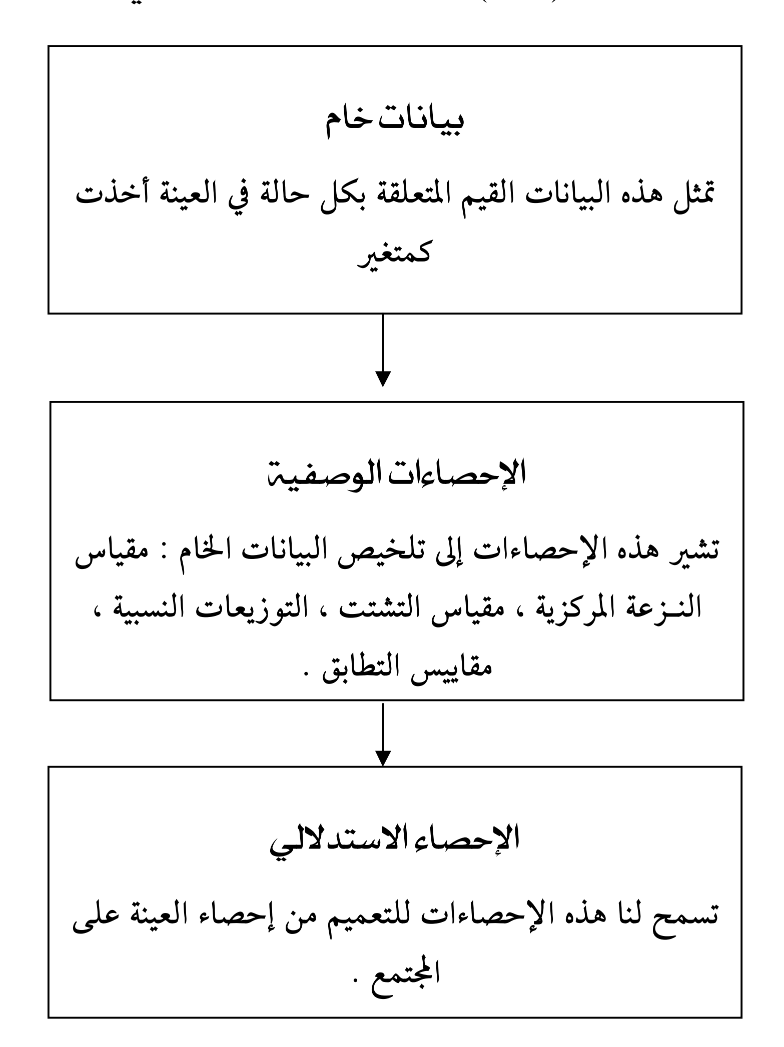
1- **بيانات خام:** وتمثل هذه البيانات المقاييس التي أخذت من كل حالة للمتغير. فالعمر على سبيل المثال تم قياسه بعدد السنوات. إن مثل هذه البيانات عادة ما تكون بيانات كثيرة اعتماداً على حجم العينة الفعلي.

2- **إحصاءات العينة:** وهذه تشير إلى الإحصاءات الوصفية التي تلخص البيانات الخام التي تم الحصول عليها من العينة (المتوسط، الانحراف المعياري أو التوزيع التكراري).

3- **الإحصاء الاستدلالي:** يساعدنا هذا النوع من الإحصاء للوصول إلى قرار حول

الخصائص المتعلقة بالمجتمع استناداً على إحصاءات العينة. بالرغم من التباين في تفاصيل الخطوات المتعلقة بإجراء الاستدلال من موقف إلى موقف آخر. إلا أننا نستخدم نفس الإجراء العام الذي يضمن توليد المجموعات العددية الثلاث. الشكل التالي يوضح هذا الإجراء.

شكل (10ـ1) عملية التحليل الاستدلالي



المصدر: George Argyrous , op.cit , p. 233

العينات العشوائية:

إن أهم مطلب يمكن مراعاته إذا أراد الباحث التعامل مع الإحصاءات الاستدلالية للتعميم من العينة على المجتمع هو أن يختار عينة عشوائية من المجتمع، فالاختيار العشوائي هو أسلوب المعاينة الذي يتيح لكل فرد في المجتمع أن تكون لديه نفس فرصة الاختيار ليكون عضواً في العينة. فالمسح عن طريق الهاتف على سبيل المثال قد لا يمثل البتة المعاينة العشوائية، باعتبار أن كثيراً من الناس لا يملكون جهاز الهاتف أثناء إجراء عملية المسح؛ وبالتالي فإن المسح لا يتضمن تلك الأسر التي لا تمتلك جهاز هاتف متيحة الفرصة لتلك الأسر التي تمتلك أكثر من جهاز هاتف التواجد ضمن أفراد العينة، ومن هنا قد يقع الباحث في عملية التحيز أي اختياره مجموعة من الناس بدلاً من أخرى.

المعاينة العشوائية الطبقية:

في كثير من الأحوال قد يلجأ الباحث - وفقاً لأسباب معينة - إلى التعامل مع العينات العشوائية الطبقية بدلاً من العينة العشوائية البسيطة، فالعينة العشوائية الطبقية تستخدم في مجتمع قابل للتقسيم الطبقي فكل طبقة هي جزء من المجتمع نشك في أنه مجتمع متماثل في إطار المتغير الذي نسعى للتعامل معه. ومن مميزات العينة الطبقية أنها تعطي درجة كبيرة من التمثيل من خلال تقليلها من احتمالية خطأ المعاينة.

تجدر الإشارة إلى أن أخطاء المعاينة يمكن التقليل منها من خلال عاملين أولهما: أن العينة الكبيرة تولد قليلاً خطأ معاينةٍ إذا ما قورنت بالعينات الصغيرة. كلما كان المجتمع متجانساً قَّلَّ خطأ المعاينة والعكس بالعكس. فإذا وافق 99 % من السكان حول قضية معينة فهذا يعني أن خطأ المعاينة في مثل هذه الحالة سيكون قليلاً. أما إذا تمت الموافقة من قِبل السكان لـِ 50 % مقابل 50 % حول القضية المطروحة ففي هذه الحالة نتوقع خطأً معاينةٍ أكبرَ. ومن هنا نجد أن العينة الطبقية تستند في الأساس على العامل الثاني من نظرية المعاينة، ولكي نتحصل على عينة ممثلة لطلاب الجامعة على سبيل المثال، ينبغي أولاً أن نتحصل على قائمة للطلاب من خلال الكليات، وبعد ذلك يمكننا سحب عينات ملائمة من الطلاب حسب السنوات أو حسب السنوات وحسب التخصص.

أما في العينة غير الطبقية فإن التمثيل من خلال الكليات فقط سيؤدي بالضرورة إلى خطأ المعاينة[1] وعادة ما نطلق على المعاينة العشوائية المعاينة الاحتمالية. وتجدر الإشارة هنا إلى أن هناك عدداً كبيراً من تقنيات المعاينة غير الاحتمالية (ليست عشوائية) وتعتبر معاينة الكرة الثلجية شكلاً من أشكال المعاينة العرضية. وهذا النوع من المعاينة يكون ملائماً عندما يكون هناك عددُ من الناس يصعب على الباحث تحديدهم أي مجتمعُ غيرُ موحدِ المعالمِ من حيث الحجم، والتركيبة. كذلك تزداد أهمية هذه التقنية عندما يكون لدينا مجتمعُ مغلقُ، وقوي الارتباط يصعب على الباحث الدخول فيه. مثل: فئة المشردين والمهاجرين، والمهاجرين غير الشرعيين إلى آخر ذلك. ومن خلال هذه التقنية يقوم الباحث بجمع البيانات من مجموعة محددة من المبحوثين التي تمكن الباحث من الوصول إليها وعندما يقوم الباحث بطرح مجموعة من الأسئلة عليهم يطلب منهم الإدلاء بأشخاص آخرين يعرفونهم. إذاً معاينة الكرة الثلجية تشير إلى عملية تراكمية حيث إن كل شخص يتم مقابلته يمكن أن يشير إلى شخص أو أشخاص آخرين يعرفهم. ولما كانت المعاينة التلجية معاينة غير احتمالية فإنها أسلوب يمكن استخدامه بشكل فعال في الدراسات الاستكشافية[2].

وتجدر الإشارة في هذا السياق إلى أنه لا يوجد لدينا أي مبرر متأصل يجعلنا نفكر في أن المعاينة الاحتمالية هي أفضل من المعاينة غير الاحتمالية باعتبار أن كل طريقة من هذه الطرق تكون ملائمة لتساؤلات بحثية مختلفة. وفي بعض الأحيان يكون من الأفضل اختيار أسلوب المعاينة غير الاحتمالية للسؤال البحثي المطروح. وتجدر الإشارة هنا، إلى أن المهتمين والأكاديميين لا ينظرون دائماً بهذه الطريقة. فالبحث يتطلب نظرة علمية عند طرحه في إطار الإحصاء الاستدلالي.

وغالباً ما يندرج في إطار العمل ليتلاءم مع التقاليد العلمية. فالإحصاء الاستدلالي في بعض الأحيان يمكن حسابه من عينة لم يتم اختيارها عشوائياً، وفي أحوال أخرى، قد يبني مشروع البحث بطريقة ما تكون الإحصاءات الاستدلالية ملائمة لهذه الخطة بالرغم من أن هناك تقنياتٍ أخرى أكثر نفاداً. إن هذه هي مشكلة مرتبطة بالممارسة البحثية التي تثير قضايا أوسع لا يمكننا التعامل معها هنا. إلا أن الشيء الذي يمكننا أن ننوه عنه في

هذا الموضع، أن اختيار تقنية البحث يجب ألا تؤخذ على أساس التقنيات التي تستخدم في تحليل البيانات، وإنما اختيار تقنيات البحث يجب أن تكون على أساس أفضل طرح للمشكلة مثار البحث.

توزيع المعاينة لإحصاء العينة:

يُطَبَّقُ الإحصاءُ الاستدلالي فقط على العينات العشوائية باعتبار أن الأداة الأساسية لإجراء عملية الاستدلال تتوقف أساساً على المعاينة العشوائية ويُطْلَق على هذه الأداة توزيعُ المعاينة لإحصاء العينة. وقبل الدخول في تعريف توزيع المعاينة دعنا نبين فكرة خلق بناء توزيعات المعاينة من خلال العملية التالية: ففي المثال المتعلق بالمجتمع المحلي الذي تم مسحه وتسجيل الأعمار بالسنوات وأن معلمات هذا المجتمع البالغ عدده 400 مفردة تكون:

$$u = 35$$

$$\sigma = 13$$

دعنا الآن نفترض أننا لم نُجْرِ مسح الحالات البالغ عددها 400 شخصاً في هذا المجتمع المحلي وعوضاً عن ذلك قمنا بإجراء التجربة التالية:

أولا: اختيار 120 حالة عشوائية وَطُرِحَ عليهم السؤالُ المتعلقُ بالعمر وتم حساب متوسط أعمارهم.

ثانيا: إعادة هذه المجموعة إلى المجتمع المحلي الأصلي مرة ثانية وسحبت مجموعة أخرى تتكون من 120 شخصاً. (ربما هذه المجموعة تحتوي على أعضاء من العينة الأولى) ونستمر في سحب هذه العملية مرة تلو الأخرى، وفي كل مرة نقوم بحساب المتوسط العمري لكل عينة تم سحبها عشوائياً. والنتيجة النهائية لعدد 20 عينة عشوائية مكررة يوضحها الجدول التالي:

جدول (10 ـ 1) توزيع عشرين متوسط لعينات عشوائية

رقم العينة	متوسط العينة
1	34.7
2	35.9
3	35.5
4	34.7
5	34.5
6	35.4
7	35.7
8	34.6
9	37.4
10	35.3
11	34.1
12	35.5
13	34.9
14	36.2
15	25.6
16	35.0
17	35.1
18	36.4
19	35.6
20	33.6

من خلال بيانات هذا الجدول يتبين لنا أن معظم المتوسطات تتجمع حول قيمة المجتمع وهي 35 سنة مع وجود درجات قليلة تبتعد بعض الشيء عن قيمة متوسط المجتمع، وهي الدرجة المتطرفة 37.4 سنة. وقد جاء هذا المتوسط لهذه العينة عن طريق الصدفة خلال عملية التباين العشوائي لتشمل نسبياً مجموعة من الأعمار الكبيرة في المجتمع المحلي. وبالرغم من ذلك يجدر بنا أن نورد الحقيقة التي مفادها أن أعمار الأشخاص في هذا المجتمع تتراوح ما بين 2 و 69 سنة وبالتالي فإن متوسط العينات تحتوي على مدًى محدودٍ من القيم قرابة نصف عدد العينات العشرين التي تم سحبها والتي تولد متوسطاً عمريّاً داخل حدود نصف سنة من المتوسط الحقيقي للمجتمع وهذا يعطينا معنى بقيمة وثبات العينات العشوائية.

دعنا نتعمق أكثر في هذا المثال النظري ونتخيل نظرياً أننا قد أخذنا عدداً لا متناهي من العينات العشوائية متساوية في الحجم من هذا المجتمع ثم نلاحظ التوزيع لكل المتوسطات المتعلقة بهذه العينات. إن النمط الذي قمنا بمشاهدته أعلاه فإننا قد لاحظنا فقط أن عشرين عينة عشوائية سوف تُعزز. أن معظم العينات تتركز حول معلمات المجتمع مع وجود بعض العينات نسبياً تبتعد عن التوزيعات بشكل أو آخر. إن مثل هذا التوزيع يطلق عليه توزيع المعاينة. وتُعْرَفُ توزيعات المعاينة بأنها التوزيعات الاحتمالية النظرية الناتجة عن عدد لا متناهي من نتائج إحصاء العينة باستخدام عينات عشوائية متساوية الحجم، بمعنى آخر أن توزيعات المعاينة هي توزيعات نظرية تعتمد في بنائها على الممارسة المنطقية. كما يُعْرَفُ توزيع المعاينة: بأنها التوزيعات التي تصف ظهور قيم التوزيعات لبعض الإحصاء تم حسابها من كل أحجام العينات التي تم سحبها من بعض المجتمعات. وإذا ما نظرنا إلى مثل هذه التعريفات بشكل دقيق يمكننا القول:

1- توزيع المعاينة هو توزيع نظري لكينونات entities. ومع أن توزيعات المعاينة نظرياً ممكنة لسحب كل العينات الممكنة من المجتمع إلا أنه قد يتعذر من الناحية العملية على الأقل إذا كان حجم المجتمع كبيراً.

2- توزيعات المعاينة لا تكشف عن ظهور الدرجات التكرارية على بعض من المتغير ولكنها تصف بدلاً من ذلك لماذا في أحوال كثيرة نجد أن بعض القيم المتنوعة لبعض

الإحصاء تم الحصول عليه من خلال حساب العينات التي تم سحبها من المجتمع. على سبيل المثال. إن توزيعات المعاينة للمتوسط يتشكل من خلال سحب كل الأحجام الممكنة للعينة (N) من المجتمع وحساب المتوسط لكل عينة. كما يمكننا أيضاً أن نتخيل أننا قمنا بسحب عدد من الأحجام الممكنة للعينات (N) من المجتمع وقمنا بحساب التباين لكل عينة وهذا ما يطلق عليه توزيعات المعاينة للتباين.

3- إن توزيعات المعاينة ليست دائماً تستند على سحب كل العينات الممكنة من المجتمع فبعض توزيعات المعاينة يمكن الحصول عليها من خلال سحب كل الأزواج الممكنة من العينات، عينات ثلاثية أو عينات رباعية أو أي مجموعة متعددة من العينات ويحسب الإحصاء لكل زوج من العينات أو العينات الثلاثية أو الرباعية أو أي مجموعة متعددة من العينات تشكل بذلك هذه الحسابات الإحصائية توزيعات المعاينة[3].

إن توزيعات المعاينة للمتوسطات لها خصائص ثلاثة مهمة:

- إن متوسط توزيع المعاينة يكون مساوياً لمتوسط المجتمع بمعنى آخر أن متوسط المتوسطات ($u\overline{x}$) سيكون نفسه مثل متوسط المجتمع ويمكن التعبير عنه جبرياً بـ:

$$u\overline{x} = u$$

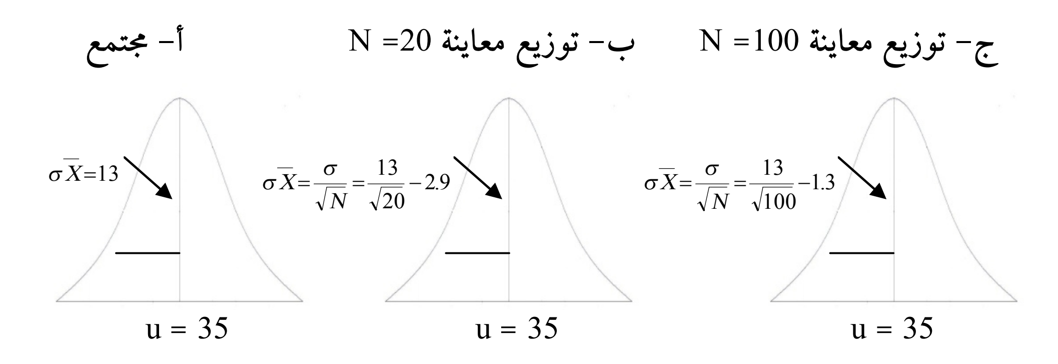
إن الخطأ المعياري سوف يرتبط بالانحراف المعياري للمجتمع. فالانحراف المعياري لتوزيع المعاينة يُعْرَفُ بالخطأ المعياري ($\sigma\overline{x}$) Standard error. وأن قيمته تتأثر بحجم العينة وكمية التباين في المجتمع. فإذا أخذنا على سبيل المثال العينة من خمسة أشخاص. وأن واحداً من هؤلاء الأشخاص في هذه العينة الصغيرة يبلغ من العمر 60 عاماً. فإن متوسط هذه العينة سيكون متأثراً بشكل كبير بهذه الدرجة. بمعنى آخر، أننا نتوقع أن تكون هذه العينة الصغيرة أقل ثباتاً إذا ما قورنت بالعينة الكبيرة، حيث من المتوقع أن تكون هناك احتمالية كبيرة للحصول على نتيجة واسعة من التشتت. أما إذا كان لدينا عينة حجمها 220 مفردة. فإن تأثير إحدى الدرجات الكبيرة سيضعف بتأثير مجموعة كبيرة من الحالات التي تكون قريبة من متوسط المجتمع. وعليه فإن إعادة تكرار عينات كبيرة ستتجمع قريباً من قيمة المجتمع؛ وبالتالي ستكون هذه العينات أكثر ثباتاً. وبشكل

مماثل إذا قمنا بسحب عينات من مجتمع تتوزع الأعمار فيه من سنتين إلى مئة واثنتين سنة، فإن مدى الدرجات التي سنتحصل عليها من هذه العينات ستكون أكبر مما لو قمنا بالمعاينة من مجتمع تتوزع فيه الأعمار بين 20 سنة و 30 سنة. إذاً كلما كان المجتمع متجانساً تجمعت العينات العشوائية بشكل وثيق حول ذلك المجتمع الذي سحبت منه هذه العينات. هذان العاملان يجذبان الانتباه من خلال المعادلة التالية للخطأ المعياري:

$$\sigma \overline{X} = \frac{\sigma}{\sqrt{\ }}$$

إن توزيعات المعاينة ستتوزع توزيعاً طبيعياً وأن نسبة العينات ستقع داخل مدى محدد من القيم التي ستحدد من خلال التوزيع الطبيعي المعياري.

إن الخصائص الثلاثة لتوزيعات المعاينة لمتوسطات العينة يوضحها الشكل التالي:



شكل (10 ـ 2) توزيعات المعاينة لعينات ذات أحجام مختلفة

من الشكل (10 – 2) يتضح أن شكل (أ) يبين توزيع الحجم الكلي للمجتمع المحلي البالغ عدده 400 نسمة. أما شكل (ب) فيشير إلى توزيع المعاينة لمتوسطات عينات يبلغ حجمها 20 مفردة. بمعنى آخر، هو توزيع المتوسطات التي ستتحصل عليها إذا ما كررنا سحب 20 مفردة من سكان هذا المجتمع. أما الشكل (ج) فهو توزيع المعاينة لمتوسطات عينة حجمها 100. من خلال هذه الأشكال يمكننا أن نرى أن كلا التوزيعين يتمركزان

عند متوسط المجتمع 35 سنة إضافة إلى أنهما موزعان توزيعاً طبيعياً، ومع ذلك فإن الخطأ المعياري لكل واحد من توزيعات المعاينة سيتباين. ومع تكرار عينات حجمها 20 فإن متوسطاتها ستتوزع بشكل كبير بخطأ معياري يصل إلى 2.9 سنة في حين أنه في العينات الكبيرة عندما كان حجم العينة 100 فإن نتائج هذه العينة تتجمع بشكل وثيق حول قيمة المجتمع وكلا التوزيعين موزعان توزيعاً طبيعياً، وبالتالي نجد أن 68 % لكل الحالات تقع خلال (1) انحراف معياري من المتوسط. في حين أن توزيع المعاينة عندما يكون حجم العينة 20 فإن هذا المدى سيكون بين 32.1 سنة و 37.9 سنة.

$$35 \pm 2.9 =$$

32.1 سنة

و

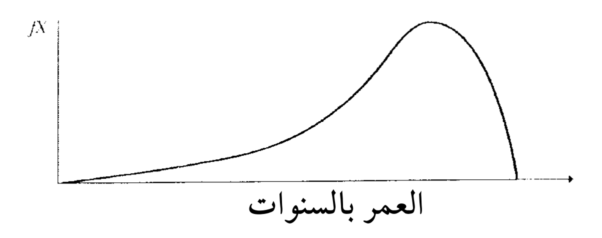
37.9 سنة

في حين نجد أن المدى في العينة الثانية لتوزيع المعاينة سيكون أكثر ضيقاً من توزيع المعاينة للعينة الأولى فالحد الأدنى 33.7 والحد الأعلى 36.3 سنة [4].

نظرية تقارب التوزيعات الاحتمالية:

قد تمت مناقشة خصائص توزيع المعاينة مستنتجة من مجتمع موزع توزيعاً طبيعياً. وعلى وجه الخصوص، فإن توزيع المعاينة هي الأخرى موزعة توزيعاً طبيعياً.

السؤال الذي يمكن طرحه هنا هو: ماذا لو أن أعمار الـ 400 نسمة لهذا المجتمع المحلي الصغير قد توزعت أعمارهم وفقاً للشكل التالي:



شكل (10 ـ 3) توزيع مائل

إن هذا التوزيع هو توزيع مائل نحو اليسار مشيراً إلى أن هناك نسبياً عدداً أكبر من كبار السن مقارنة بصغار السن بهذا المجتمع المحلي. وأن تكرار عينات عشوائية من هذا التوزيع المائل سيولد لنا كذلك. توزيعات معاينة مائلة أيضاً. ومع هذا فإن الأمر ليس كذلك. وطبقاً لإحدى القواعد الإحصائية الرئيسية: نظرية تقارب التوزيعات الاحتمالية Central Limit Theorem، أوضحت أن ثمة شروطاً معينة ستكون فيها توزيعات المعاينة مُوَزَّعُ توزيعاً طبيعياً، بالرغم من أن توزيع المجتمع الذي سحبت منه العينة موزعاً توزيعاً غير طبيعي. وتعرف نظرية تقارب التوزيعات الاحتمالية، بأنها عدد غير محدود من العينات العشوائية بأحجام متساوية تم اختيارها من مجتمع، وأن توزيع المعاينة لمتوسطات العينة سيقترب من التوزيع الطبيعي مثلما يقرب حجم العينة من اللامتناهي.

إن المجتمع يمكن أن تكون وحداته موزعة توزيعاً غير طبيعي، ومع ذلك فإن تكرار المعاينة (نظرياً) سيولد توزيعات معاينة طبيعية. في حقيقة الأمر، إن حجم العينة ليس بالضرورة أن يكون كبيراً كما تم افتراضه في الصيغة الأساسية لنظرية تقارب التوزيعات الاحتمالية: فعندما يزيد حجم العينة عن 100، فإن توزيع المعاينة لمتوسطات العينة سيكون على وجه التقريب طبيعياً [5].

خلاصة القول، بالرغم من النقاش المتعلق بتوزيعات المعاينة إلا أنه من الناحية العملية في مجال البحوث الاجتماعية والإنسانية عادة ما يلجأ الباحث إلى سحب عينة واحدة من مجتمع ومن ثم يقرر هل هذه العينة تعكس الخصائص العامة للمجتمع أم لا. إن السؤال الذي يمكن طرحه هنا هو: ما الفائدة من استخدام توزيعات المعاينة؟ إن توزيعات المعاينة تعتبر الحجر الأساس الذي يستند عليه الإحصاء الاستدلالي كما سنرى في الفصول اللاحقة.

ثانياً: التقدير وفترات الثقة:

في الجزء الأول من هذا الفصل حاولنا أن نبين الخصائص المتعلقة بتوزيع المعاينة لمتوسطات العينة. حيث يتميز توزيع المعاينة بثلاث خصائص مهمة:

1- إن متوسط توزيع المعاينة يكون مساوياً لمتوسط المجتمع، بالرغم من أن متوسط أي

عينة يمكن أن يختلف عن المجتمع الذي سحبت منه هذه العينة، وأن متوسطات العينة العشوائية المكررة سوف تتمحور حول القيمة "الحقيقية" للمجتمع.

$$u\overline{x} = u$$

بمعنى آخر، بالرغم من أن النتائج الفردية التي سوف تتباين من عينة إلى عينة أخرى في معدل متوسطات العينة ستكون مساوية لذلك المعدل في المجتمع. فإن هذه الخاصية المتعلقة بمتوسط العينة تجعل منه تقديراً غير متحيز لقيمة المجتمع. فإحصاء العينة إحصاء غير متحيز إذا كان توزيع المعاينة لديها متوسطاً مساوياً لمعلمات المجتمع يكون إحصاء العينة قادراً على تقدير هذه المعلمة.

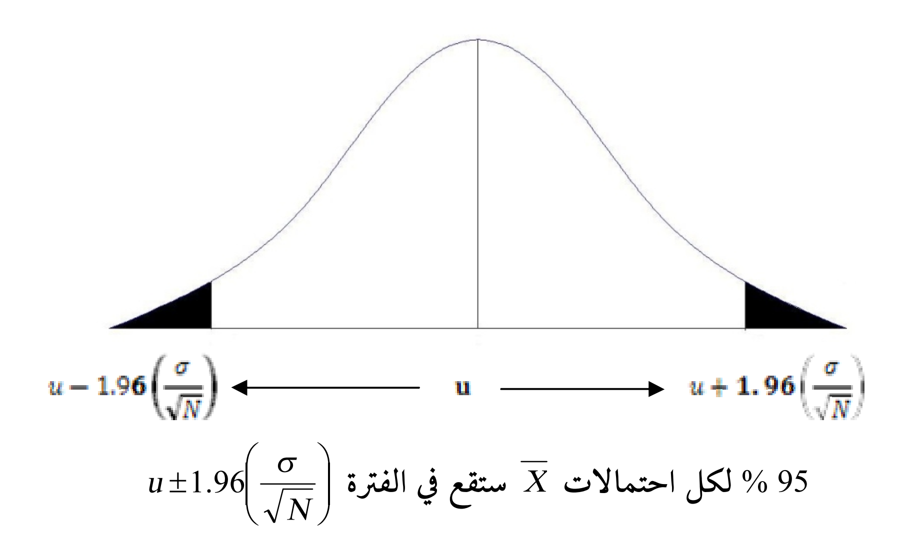
2- إن انتشار نتائج العينة حول قيمة المجتمع يكون متأثراً بحجم العينة. فالانحراف المعياري لتوزيع المعاينة، يطلق عليه الخطأ المعياري الذي يعرف بالمعادلة التالية:

$$\sigma\overline{X} = \frac{\sigma}{\sqrt{}}$$

وبزيادة حجم العينة فإن الخطأ المعياري لتوزيع المعاينة يصبح أصغر، وبذلك فإن نتائج العينة تكون أكثر إحكاماً في التمحور حول قيمة المجتمع. بمعنى آخر، إن العينات الكبيرة تمدنا بتقديرات فعالة لقيم المجتمع.

3- توزيعات المعاينة لمتوسطات العينة تكون طبيعية، تناسب متوسطات العينة وستقع داخل مدى محدد من القيم التي ستحدد عبر التوزيع الطبيعي المعياري. ومن خلال هذه الحقيقة المحددة يمكننا استخدام التقنيات التي تم عرضها في الفصل المتعلق بالمنحنى الطبيعي المعياري الذي يتم وفقاً له استخدام درجات Z لوصف توزيع الحالات عبر مدى من القيم.

إن هذه الخاصيات الثلاثة لتوزيع المعاينة لمتوسطات العينة تمكننا من مراجعة جدول المساحة تحت المنحنى الطبيعي المعياري (انظر الملحق المتعلق بالمساحة تحت المنحنى الطبيعي) من أجل ضمان الاحتمالية بأن أي متوسط عينة معطى سيكون داخل مدى محدد من القيم حول متوسط المجتمع. على تقدير أن حوالي 95 % من العينات المكررة سيكون لها متوسط داخل 1.96 أخطاء معيارية عن متوسط المجتمع (انظر الشكل التالي):



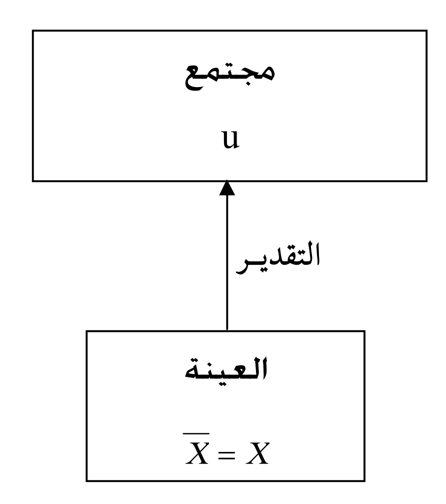
شكل (10-4) توزيع المعاينة لمتوسطات عينة

الجدير بالذكر، أن هذا التوزيع يسمح لنا بتحديد مدى أو درجات الفترة داخل 95% لكل متوسطات العينة المحتملة، وسوف تقع داخلها. وتعرف بالمعادلة التالية:

$$u \pm 1.96\left(\frac{\sigma}{\sqrt{N}}\right)$$

التقدير:

في الشكل (10 - 3) تم طرح المشكلة بطريقة معينة، فقد كانت لدينا معلمة مجتمع، وتم تقدير مدى القيم وهي 95 % لكل العينات العشوائية التي تم سحبها من ذلك المجتمع الذي نتعامل معه وعلى أية حال في القضايا البحثية عادة ما تُطْرَحُ المشكلة نفسها بشكل آخر، فإذا كان لدينا نتيجة لعينة واحدة ونحتاج إلى تقدير قيمة المجتمع من العينة فإن ذلك يمكن توضيحه (من خلال الشكل 10 - 5).



شكل (10 ـ 5) عملية تقدير متوسط المجتمع

في الجزء الأول من هذا الفصل تم تحليل 400 شخصٍ في مجتمع محلي افتراضي. وقد كان متوسط هذا المجتمع المحلي 35 سنة، وسحبت عينة عشوائية منه بحجم 120 شخصٍ N = 120، وقد تبين لنا أن المتوسطات لكل واحد من هذه العينات ليست مساوية لمعلمة المجتمع الذي سحبت منه هذه العينات، إلا أنه يمكننا القول بأن معظم هذه المتوسطات تتجمع حول قيمة المجتمع. والسؤال الذي يمكن طرحه هنا هو: ما الأمر إذا لم يكن لدينا معرفة بأن المتوسط العمري لهذا المجتمع كان 35 سنة، وأن كل ما في وسعنا أن نتعامل مع واحدة من هذه العينات البالغ عددها 120 مفردة؟ دعنا نفترض أن واحدة من هذه العينات وهي العينة الوحيدة التي ولدت متوسط عمري يصل إلى 34.5 سنة وأن المهمة التي تقع على عاتقنا هي تقدير معلمة المجتمع (لاحظ هنا أننا نتظاهر بأنه ليس لدينا أية معلومة حول معلمة المجتمع) من نتيجة عينة واحدة. ولكي تقدر معلمة المجتمع ينبغي علينا أن تبدأ بفرضية. وتقترح هذه الفرضية أن العينة تقع فعلياً داخل منطقة معينة في توزيع المعاينة. ونفترض أن متوسط العينة ليس واحداً من هذه المتوسطات القليلة بعيدة الاحتمال. والنتائج المتطرفة تكون مختلفة كثيراً عن قيمة المجتمع. فعلى سبيل المثال، قد ينتابنا شعور بالراحة مع الفرضية التي مفادها: أن هذه العينة المكونة من 120 شخصاً تكون واحدة من 95 % من كل العينات المحتملة التي تقع داخل 1.96 ± انحرافات

معيارية من متوسط المجتمع. ينبغي علينا أن نضع في أذهاننا أن هذا الأمر مجرد فرضية: فقد يكون في حقيقة الأمر أننا قد سحبنا واحدة من هذه العينات القريبة بمتوسط يختلف كثيراً عن معلمة المجتمع. وأنه ليس باستطاعتنا البتة معرفة ما إذا كان الأمر كذلك، ولكن مع وجود احتمالية قليلة جداً قد يكون الأمر كذلك (أقل من 5 في 100)، فالفرضية هنا تبدو معقولة. بمعنى آخر، أننا سنكون على ثقة بأن هذه الفرضية تكون صحيحة. ومن هنا نطلق على هذا الأمر، الاحتمالية المفترضة لمستوى الثقة؛ في هذا المثال، قد تعاملنا مع مستوى ثقة 95 %.

ومع الأخذ بهذه الفرضية فإن نتيجة العينة تكون داخل مدى 95% لكل نتائج العينة المحتملة ومن ثم نكون قادرين على تقدير المجتمع. نحن نعلم أن توزيع المعاينة سيكون لديه انحراف معياري يطلق عليه الخطأ المعياري:

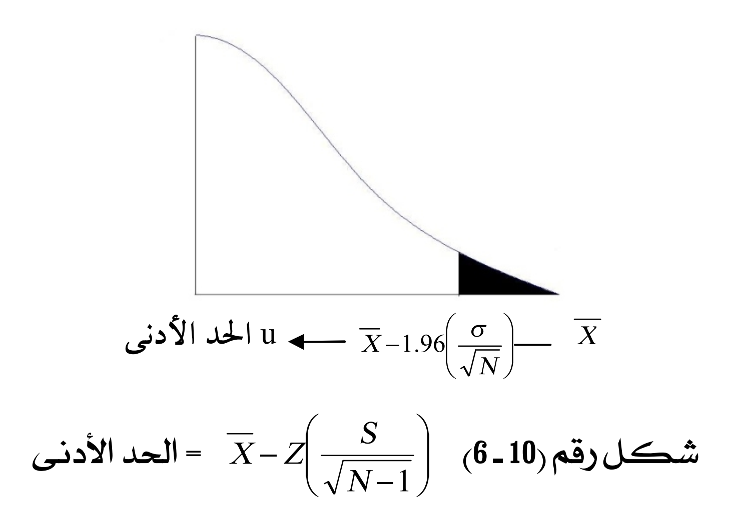
$$\sigma \overline{X} = \frac{\sigma}{\sqrt{\ }}$$

وعندما لم يكن لدينا أي معرفة بالانحراف المعياري للمجتمع ( ) لذلك يمكننا استخدام الانحراف المعياري للعينة بدلاً من ذلك. والذي سيكون في هذه العينة مساوياً لـِ 13 سنة:

$$\sigma \overline{X} = \frac{S}{\sqrt{N-1}} = \frac{13}{\sqrt{120-1}}$$

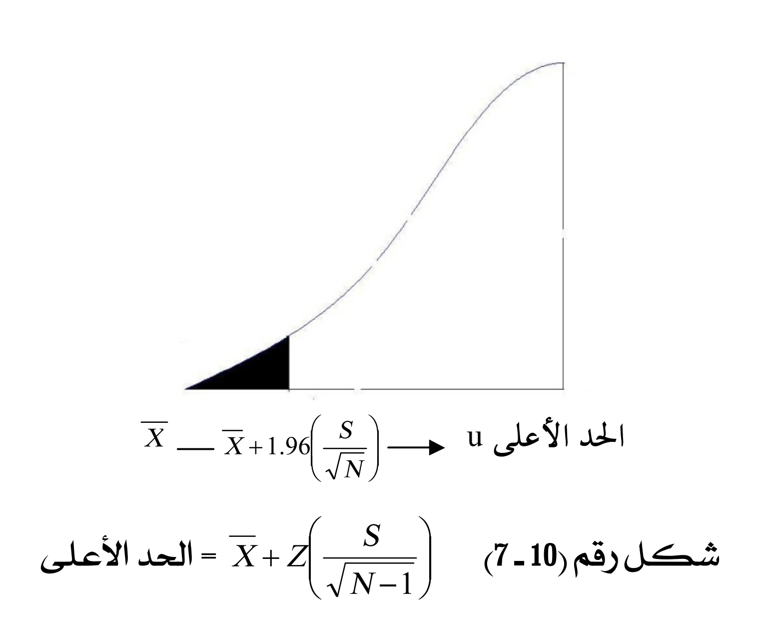
$$= 1.2 \text{ سنة}$$

ومع الأخذ بهذه القيمة للخطأ المعياري فإن أبعد معلمة للمجتمع ستكون تحت قيمة العينة وأن هذه القيمة أي قيمة العينة تبقى داخل منطقة 95 % (1.96-) انحرافات معيارية. ونطلق على ذلك الحد الأدنى Lower Limit للتقدير. ويحدد الحد الأدنى للتقدير المسافة القصوى التي ستكون فيها قيمة المجتمع تحت قيمة العينة. انظر شكل (6).



شكل رقم (10 ـ 6) $\overline{X}-Z\left(\frac{S}{\sqrt{N-1}}\right)$ = الحد الأدنى

وباستخدامنا لنفس المنطق، فإن أبعد معلمة للمجتمع يمكن أن تكون أعلى من قيمة العينة وتكون داخل منطقة 95 %، 1.96 + أخطاء معيارية ونطلق على هذا الحد الأعلى Upper Limit كما هو موضح في الشكل (10 – 7).



شكل رقم (10 ـ 7) $\overline{X}+Z\left(\frac{S}{\sqrt{N-1}}\right)$ = الحد الأعلى

وبوضع هذين الجزئين من المنطق معاً يسمحان لنا بتعريف مدى القيم التي يطلق عليها فترة الثقة Confidence Interval (C.i)، ونعني بفترة الثقة مدى القيم التي يمكن تقديرها، وتشمل إحصاء المجتمع على مستوى محدد من الثقة.

إن خطوات تحديد الحد الأدنى والحد الأعلى لفترة الثقة يمكن جمعها في المعادلة التالية:

$$Ci = \overline{X} \pm Z\left(\frac{S}{\sqrt{N-1}}\right)$$

حيث إن: c.i فترة الثقة.

$\overline{X}$ متوسط العينة.

Z درجة Z كما تحدد من خلال ألفا.

$\frac{S}{\sqrt{N-1}}$ الانحراف المعياري للعينة.

$\frac{\sigma}{\sqrt{N}}$ الانحراف المعياري للمجتمع.

$$u \leftarrow \overline{X} - 1.96\left(\frac{\sigma}{\sqrt{N}}\right) — \overline{X} — \overline{X} + 1.96\left(\frac{\sigma}{\sqrt{N}}\right) \rightarrow u$$

Lower Limit الحد الأدنى $\quad Ci = \overline{X} \pm 1.96\left(\frac{\sigma}{\sqrt{N}}\right) \quad$ Upper Limit الحد الأعلى

تبين لنا هذه المعادلة أنه يمكن إضافة أو طرح مسافة من نتائج العينة عرفت بالرقم الأقصى لدرجات Z التي افترضنا أن نتيجة العينة جاءت من معلمة المجتمع. إننا نتعامل مع درجة Z لـِ 1.96. لأننا افترضنا أن العينة كانت واحد من 95 % من كل العينات الممكنة التي تقع قريبة من متوسطة المجتمع. ففي المثال الذي نتعامل معه، فإن الحد الأدنى والحد الأعلى لأولئك القاطنين في هذا المجتمع المحلي يكون:

$$\text{الحد الأدنى} = \overline{X} - Z\left(\frac{S}{\sqrt{N-1}}\right) = 34.5 - 1.96\left(\frac{13}{\sqrt{120-1}}\right)$$

$$= 32.3 \text{ سنة}$$

$$\text{الحد الأدنى} = \overline{X}+Z\left(\frac{S}{\sqrt{N-1}}\right)=34.5+1.96\left(\frac{13}{\sqrt{120-1}}\right)$$

= 36.8 دقيقة

ويمكن التعبير عن مثل هذا التقدير بالشكل التالي:

[36.8 و 32.3] 34.5. قد تم بناء فترة ثقة لأن تقدير المتوسط العمري للمجتمع من عينة واحدة يحتاج منا أن نعطي مساحة لتأثير تباين المعاينة. وبالنظر إلى التقدير الذي تم بناؤه من هذه العينة، فإنه بإمكاننا أن ندرك أن هذا التقدير يتضمن المتوسط الفعلي للمجتمع البالغ 35 سنة. والذي أدعينا أننا لا نعرفه. إن فترة الثقة تكون دقيقة، ذلك أن مدى القيم يتراوح بين 32.3 و 36.8 سنة. وهذا المدى من القيم يحتوي المتوسط الفعلي للمجتمع. والجدير بالذكر أنه في الحالة السوية، لا نعرف ما إذا كان التقدير دقيق، سوى أن مستوى الثقة يشير إلى احتمالية كونه دقيق [6].

تغير مستوى الثقة:

من خلال المناقشة أعلاه، اخترنا مستوى ثقة 95 % وهذا ما حدا بنا إلى مضاعفة الخطأ المعياري لدرجة Z بـِ 1.96، نظراً لأن هذا يحدد المساحة تحت توزيع المعاينة التي تحتوي 95 % من نتائج العينة المكررة. إن مستوى ثقة 95 % هي الأكثر شيوعاً واستخداماً، إلا أن ذلك لا يمنع البتة من أن يختار الباحث إما مستوى أكبر أو أصغر استناداً إلى كم من الثقة يريد الباحث اعتمادها. ونلاحظ أنه كلما كان مستوى الثقة أكبر فمن المرجح أن فترة الثقة التي تشتق منها ستتضمن متوسط المجتمع والعكس بالعكس. إذا ما اخترنا على سبيل المثال، فترة ثقة 99 % إذاً نحن نفترض أن متوسط عينة محددة يكون 1 من 99 في 100 سيقع 2.58 انحرافات معيارية في أي اتجاه للمتوسط الحقيقي:

$$Ci=\overline{X}\pm 2.58\left(\frac{S}{\sqrt{N-1}}\right)$$

ولزيادة التوضيح لفهم التأثير الحاصل من اختيار مستويات ثقة مختلفة، دعنا نوضح ذلك من خلال المثال التالي:

مثال: عينة عشوائية مكونة من 200 فردٍ تم سحبها. وطرح سؤال على كل واحد من هؤلاء حول دخله السنوي، وقد وصل متوسط دخل هذه العينة سنوياً إلى 35.000 دولار بانحراف معياري 5000 دولار:

$$\overline{X} = 35.000$$

$$S = 5000$$

$$N = 200$$

السؤال المطروح نطرحه هنا: ما هو تقديرنا لمتوسط الدخل السنوي لهؤلاء الناس بمستوى ثقة 95 %؟

$$Ci = \overline{X} \pm Z\left(\frac{S}{\sqrt{N-1}}\right) = 35.000 \pm 1.96\left(\frac{5000}{\sqrt{200-1}}\right)$$

$$= 35.000 \pm 695$$

الحد الأدنى والحد الأعلى هما

الحد الأدنى للفترة هو $35.000 - 695 = 34.305$

الحد الأعلى للفترة هو $35.000 + 695 = 35.695$

وبذلك فإن تقديرنا لمتوسط الدخل السنوي لكل هؤلاء الناس بمستوى ثقة 95 % يقع داخل المدى التالي:

$$34.305 \leq u \leq 35.695$$

إذاً فاتساع أو عرض الفترة: $35.695-34.305=1390$ = اتساع أو عرض الفترة (Interval Width)

ويمكننا أن نتتبع نفس الإجراء إذا أردنا أن نولد فترة ثقة 99 %

$$Ci = \overline{X} \pm Z\left(\frac{S}{\sqrt{N-1}}\right) = 35.000 \pm 2.58\left(\frac{5000}{\sqrt{200-1}}\right)$$

$$= 35.000 \pm 915$$

$$= 35.000\ [34.085 \text{،} 35.915]$$

ولكي نكون أكثر ثقة بأن فترة الثقة سوف تحتوي فعلياً القيمة الحقيقية للمجتمع، وتصبح أكثر اتساعاً؛ فهي تتراوح الآن في هذا المثال، من 34.085 إلى 35.915 بعرض يبلغ 1830 (35.915–34,085=1830). إذا ما أردنا على الجانب الآخر، أن نكون أكثر دقة في تقديرنا فإنه بالإمكان اختيار مستوى ثقة 90 %، غير أن هذا سيكون فيه مخاطرة كبيرة على أن نكون مخطئين في التقدير، وبالتالي فإن فترة الثقة لهذا المستوى لدرجة Z تساوي 1.645:

$$Ci = 35.000 \pm 1.645\left(\frac{5000}{\sqrt{200-1}}\right)$$

$$= 35.000\ [34.415 \text{،} 35.585]$$

إن تأثير التغيرات في مستوى الثقة في تقديراتنا يمكن تلخيصه في الجدول التالي:

جدول (10 ـ 2) تأثير مستوى الثقة على فترة الثقة (N= 200)

مستوى الثقة ألفا (∝) %	درجة Z	فترة الثقة	مدى الفترة
% 90	1.65	35.000 ± 585	1170
% 95	1.96	35.000 ± 695	1390
% 99	2.58	35.000 ± 915	1830

المصدر: George Argyrous, op.cit , p. 254

نلاحظ أنه باستخدامنا لمستوى ثقة أصغر يقلل من اتساع أو عرض الفترة التي قدرنا فيها قيم المجتمع. وبالرغم من أن مدى هذه الفترة يكون صغيراً فإن فرصة وقوعنا في الخطأ (طبقاً لمستوى الدلالة) سيزداد أيضاً وبتطبيق مدى القيم تزيد الفرصة بأنها لن

تحتوي على متوسط المجتمع. وباختيارنا لمستوى ثقة 99 % يوسع تقدير الفترة بحيث تصبح أكثر احتمالاً بأن تحتوي على قيمة المجتمع، إلا أنه من الممكن كنتيجة تعمل على أن يكون هذا التقدير غير ذي معنى من الناحية النظرية أو العملية. وبمعرفتنا بأن متوسط الدخل السنوي لأفراد هذه العينة يمكن أن يكون بين 34.085 و 35.915 ربما نحن فعلياً لم نصل شيء له أهمية عملية [7].

إجراء توليد عينات عشوائية مكررة باستخدام برنامج Spss:

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Data ⟸ Select ⟸ Cases

(يعطيك صندوق الحوار لِـ Select Cases)

2- اختيار عينة الحالات العشوائية وذلك بالنقر على الدائرة الصغيرة ⊙ بجانب هذا الخيار.

3- النقر على مفتاح Sample (هذه العملية تقودك إلى اختيار الحالات Random Sample Select Cases وتعطيك خيار الاختيار لنسب محددة من الحالات، أو عدد معين من الحالات. في هذا المثال اختار الباحث عدد (120 حالة).

4- انقر على الدائرة الصغيرة ⊙ بجانب Exactly.

5- اطبع 120.

6- اطبع 400 في المربع بجانب From the First.

7- انقر على Contiue.

8- انقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

حساب متوسط العينة Calculating The Sample Mean:

إن الخطوة التالية هي إعطاء الأمر SPSS لحساب المتوسط لهذه العينة باستخدام أمر:

analyze ⟸ Statistics ⟸ Descriptive ⟸ Frequencies

(ملاحظة ينبغي على الباحث في مثل هذا الإجراء أن يختار Only فقط المتوسط Mean حتى لا يحصل على جدول تكراري وإحصاءات وصفية أخرى لكل عينة مكررة نظراً لأن هذه العملية ستولد مخرجات لا ضرورة لها في هذا الشأن.

إجراء تكرار المعاينة Repeating The Sampling Procedure:

إن إجراء هذين الأمرين بشكل متعاقب سوف يولد متوسطاً للعينات العشوائية المختارة للحالات 120. ولسحب عينة عشوائية أخرى يتطلب:

1- اختيار أمر Data⟸Select Cases والنقر مباشرة على OK (ليس بالضرورة أن نختار مرة ثانية 120 حالة فالبرنامج بشكل آلي يعيد مجموع التعليمات السابقة ويختار عينة جديدة قوامها 120 وبالطريقة المشابهة:

analyze ⟸ Statistics ⟸ Descriptive ⟸ Frequencies

وبعد ذلك النقر على OK حينها سوف يتم حساب المتوسط بدون اللجوء إلى اختيار كل الخيارات داخل هذا الأمر. وبالتالي سوف تتحصل على 20 جدولاً كلها تشبه الشكل رقم (8).

Frequencies

Statistics

Age of Respondent

N	Valid	120
	Missing	0
Mean		35.76

شكل رقم (10-8) متوسط عينة عشوائية باستخدام SPSS

إجراء فترات الثقة Confidence Intervals باستخدام Spss:

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Analyze ⇐ Descriptive ⇐ Statistics ⇐ Explore

2- اختر TV. Watched per Night من قائمة المتغيرات Variables

3- انقر على ◄ مشيراً إلى القائمة المحددة بعنوان Dependent List هذه العملية تقوم بلصق TV. Watched Pernight إلى القائمة المحددة للمتغير التابع Dependent List.

4- انقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

Descriptives

			Statistic	Std.Error
TV Watched per night in Minutes	MEAN		165.85	6.55
	95% Confidence	Lower Bound	152.14	
	Interval for mean	upper Bound	179.56	
	5% Trimmed mean		166.94	
	Median		165.00	
	Variance		857.924	
	Std.Deviation		29.29	
	Minimum		102	
	Maximu,		210	
	Rage		108	
	Inerquartile Range		42.50	
	Skewness		-.450	.512
	Kurtosis		-.303	.992

شكل رقم (10 - 9) مخرجات أمر استكشاف SPSS

Spss Explore Comnond output

تغير حجم العينة:

بمعزل عن مستوى ألفا، فإن العامل الآخر الذي يحدد مدى فترة الثقة يكمن في حجم العينة. فإذا أبقينا على مستوى ثقة 95 %، وأن الذي يتباين هو فقط حجم العينة، فإننا بذلك نلاحظ أنه يصبح اتساع أو عرض Width الفترة أصغر بزيادة حجم العينة. انظر الجدول التالي:

جدول (10 – 3) تأثير حجم العينة على مدى الفترة (α = 0.05)

حجم العينة	مدى الفترة
100	1970
200	1390
500	877
1000	620
10.000	196

المصدر: George Argyrous, op.cit , p. 255

إن الشيء الذي نلاحظه من الجدول السابق حول تأثير حجم العينة هو أن زيادة حجم العينة له تأثير كبير على اتساع أو عرض الفترة إذا ما قورن بالعينات الصغيرة. فزيادة حجم العينة من 100 إلى 200 يخفض من اتساع الفترة بـِ 580، الذي هو أكبر من 424 عندما تم زيادة حجم العينة من 1000 إلى 10.000. ولهذا السبب يرى كثير من المشتغلين بالمسوح الاجتماعية وبحوث قياسات الرأي العام، حتى عندما يتم التعميم على ملايين من السكان نجد أن العينة المناسبة تتراوح بين 1200 إلى 1400 مفردة فقط. إن هذا الحجم من العينات يضيق فترة الثقة إلى اتساع صغير نسبياً، وإذا زاد حجم العينة أكثر من ذلك قد يؤدي إلى زيادة تكلفة إجراء البحث بدون الحصول على قدر كبير من الدقة.
ولزيادة التوضيح نورد المثالين التاليين:

المثال الأول: نفترض أننا نريد أن نقدر المتوسط العمري من كل الأطفال في روضة ما، وتم اختيار عينة عشوائية لـِ 140 طفلاً بمتوسط عمري 3.75 سنة، وبانحراف معياري 0.8 سنة، ما هو تقديرنا للمتوسط العمري لأولئك الأطفال بالروضة؟

$$\overline{X} = 3.75$$

$$S = 0.8$$

$$N = 140$$

$$Z = 1.96 \ (\alpha = 0.05)$$

$$Ci = \overline{X} \pm Z\left(\frac{S}{\sqrt{N-1}}\right) = 3.75 \pm 1.96\left(\frac{0.8}{\sqrt{140-1}}\right)$$

$$= 3.75 \pm 0.13$$

وعليه فإننا نتوقع أن المتوسط العمري لكل أطفال الروضة سيكون بين 3.62 و 3.88 سنة ويمكن كتابة هذا التقرير بالشكل التالي:

3.75 سنة = [3.88، 3.62]

المثال الثاني: عينة عشوائية من 300 شخصٍ يشاهدون الإذاعة المرئية بمعدل 150 دقيقة في الليلة الواحدة بانحراف معياري 50 دقيقة. كيف يمكن تقدير متوسط المجتمع؟

إذا أخذنا مستوى ثقة 95 % فإن الحد الأدنى للفترة هو:

الحد الأدنى = $\overline{X} - Z\left(\frac{S}{\sqrt{N-1}}\right) = 150 - 1.96\left(\frac{50}{\sqrt{300-1}}\right)$

= 144.3 دقيقة

الحد الأعلى = $\overline{X} + Z\left(\frac{S}{\sqrt{N-1}}\right) = 150 + 1.96\left(\frac{50}{\sqrt{300-1}}\right)$

= 155.7 دقيقة

وعليه فإن تقديرنا لمشاهدة الإذاعة المرئية ليلاً بمستوى ثقة 95% يكون [155.7، 144.3] 150 دقيقة [(8)].

اختيار حجم العينة*:

لتعميم أي مشروع بحثي ينبغي على الباحث أن يضع في اعتباره مجموعة من العوامل، فالمعاينة قد تكون مكلفة جداً من حيث الإجراء، وإذا أخذنا في الاعتبار عملية التدريب، والسفر، وعامل الوقت الذي يمكن أن يحيط بإجراء أي مسح واسع. وبالتالي نريد أن نضمن أن يكون لدينا عينة كبيرة وكافية تولد لنا الدقة المطلوبة، ولكن حجم هذه العينة ينبغي ألاّ يكون أكبر من الضروري.

هنا يمكننا استخدام منطق التقدير لنحدد حجم العينة الصحيح الذي نسعى إليه، آخذين في الاعتبار مستوى الدقة التي نسعى إليها. فعلى سبيل المثال، قد يكون في أذهاننا اتساع فترة محدداً؛ ففي المثال السابق، فإننا قد نرغب في تقدير دخل بعض الأفراد داخل 1000 دولارٍ كمتوسط دخل لهذه المجموعة من السكان. وبالتالي فإن اتساع أو عرض الفترة قد حدد سلفاً من خلال القضية المطروحة للبحث، والآن نعمل بشكل عكسي لنقرر حجم العينة الملائم الذي سوف يولد لنا فترة ثقة لهذا الحجم. وبدون الخوض في البراهين، فإنه يمكن اشتقاق المعادلة التالية لاختيار حجم العينة الملائم من المعادلة التي استخدمناها سلفاً:

$$N = \frac{Z^2 \times \sigma^2}{\left(\frac{width}{2}\right)^2}$$

حيث إن: Width تعني اتساع أو عرض.

في هذا المثال لدينا اتساع تم تحديده سلفاً وهو 1000 دولارٍ. وإذا اخترنا مستوى ثقة 95 %، Z = 1.96. إن الجزئية الباقية لكي نستنتج حجم العينة هي الانحراف المعياري للمجتمع.

---

* لمعرفة المزيد حول طرق تحديد حجم العينة يمكنك الرجوع إلى : مصطفى خلف عبد الجواد، الإحصاء الاجتماعي: المبادئ والتطبيقات، دار السيرة للنشر والتوزيع والطباعة، عمان، 2009، ص ص 382-387.

إن هذه الجزئية هي التي تحد من استخدام هذه التقنية. ولما كان الأمر يتعلق بإجراء تصميم مشروع بحث، أكثر منه إجراء لتحليل البيانات التي تم جمعها فعلاً، فإنه على نحو مبين ليس في قدرتنا استخدام الانحراف المعياري كبديل لـِ ، باعتبار أن العينة لم تسحب بعد ! وبالتالي فإن هذا الإجراء يكون مقتصراً لتلك المواقف التي يكون الانحراف المعياري للمجتمع فيها معروفاً، ولغرض التوضيح دعنا نفترض أن الانحراف المعياري للمجتمع معروفُ وهو 5000 دولارٍ. إنه بإمكاننا أن نستعيض عن المعلومات ذات الصلة في المعادلة لكي نحدد حجم العينة المناسب:

$$N=\frac{Z^2\times\sigma^2}{\left(\frac{width}{2}\right)^2}=\frac{1.96^2\times5000^2}{\frac{(1000)^2}{2}}$$

$$=384$$

بهذا فإننا نحتاج إلى عينة حجمها 384 إذا أردنا أن نولد فترة ثقة باتساع لا يزيد عن 1000 دولارٍ. ولتوضيح هذه الفكرة مرة أخرى، دعنا نفترض أننا نرغب في أن نكون أكثر دقة: وأن لا يكون تقديرنا أوسع من 500 دولارٍ، على مستوى ثقة 95 % ولسبب ما فإننا نريد أن نكون أكثر قرباً من القيمة المحددة. إن السؤال المطروح هو: ما حجم العينة الذي ينبغي علينا أخذها؟

$$N=\frac{Z^2\times\sigma^2}{\left(\frac{width}{2}\right)^2}=\frac{1.96^2\times5000^2}{\frac{(500)^2}{2}}$$

$$=1537$$

من خلال ما تم مناقشته أعلاه لكي نكون أكثر دقة في التقدير، فإننا نحتاج إلى عينة كبيرة، باعتبار أن العينات الكبيرة لديها خطأ معياري قليل حول معلمة المجتمع الحقيقية. ولكي تقلص الدخل السنوي لأفراد هذه العينة إلى مدى أصغر يصل إلى 500 دولار فإن الأمر يتطلب أخذ عينة من 1537 [9].

أسئلة للمراجعة:

1- بين الفرق بين المعلمة وإحصائي العينة؟

2- ما هو الفرق بين الإحصاء الوصفي والإحصاء الاستدلالي؟

3- ماذا يعني التباين العشوائي؟ وكيف يؤثر هذا التباين على قدرتنا للوصول إلى التعميم من العينة على المجتمع.

4- بين ما إذا كانت الصياغات التالية صحيحة أم خطأ:

أ- الثبات في متوسطات العينة العشوائية يعتمد على حجم العينة، والتباين في المجتمع، وحجم المجتمع.

ب- متوسطات العينات العشوائية سوف تتجمع حول متوسط المجتمع.

ج- الانحراف المعياري لمتوسطات العينة العشوائية سيكون أكبر من الانحراف المعياري للمجتمع التي سحبت منه هذه العينة؟

د- توزيع متوسطات العينة سيكون طبيعياً إذا تم سحب هذه المتوسطات من مجتمع طبيعي.

5- إذا كان متوسط مجتمع طبيعي هو 40، ما هو متوسط توزيعات المعاينة لعينة 30 = N، وعينة N = 120؟

6- ماذا يعني الخطأ المعياري؟ هل سيكون مساوياً أو أكبر أو أقل من الانحراف المعياري للمجتمع؟ ولماذا؟

7- صف لنا مشروع بحث يمكن أن تستخدم فيه عمليات العينات العشوائية الطبقية؟

8- لماذا تعتبر نظرية تقارب التوزيعات الاحتمالية مهمة في البحث؟

9- ماذا يعني التقدير بفترة؟

10- اشرح ماذا يعني مستوى الثقة؟ وكيف أن التغير في مستوى الثقة يؤثر على اتساع التقدير بفترة؟

11- بين كيف أن حجم العينة يؤثر على فترة الثقة؟

12- بين كيف أن الانحراف المعياري يؤثر في اتساع فترة الثقة؟

13- من نتائج واحد من مجموعة كل عينة من الفئات التالية، قم ببناء مستوى ثقة 95 % لتقدير u (متوسط المجتمع)؟

أ-	$\overline{X}=5.2$ S = 0.7 N = 157	د-	$\overline{X}=1.020$ S = 50 N = 329
ب-	$\overline{X}=100$ S = 9 N = 620	هـ-	$\overline{X}=7.3$ S = 1.2 N = 105
ج-	$\overline{X}=20$ S = 3 N = 220	و-	$\overline{X}=33$ S = 6 N = 220

14- من البيانات التالية التي تقيس مدة الوقت بالشهور لخريجي الجامعة للحصول على أول عمل. افترض أن هذا المتغير تم توزيعه توزيعاً طبيعياً. استنتج التقدير بفترة لمجموعة هؤلاء الخريجين مستخدماً 95 % مستوى ثقة؟

الدرجة	حجم العينة	المتوسط	الانحراف المعياري
اقتصاد	45	6	2.5
اجتماع	35	4	2.0
تاريخ	40	4.5	3.0
إحصاء	60	3	1.5

15- دراسة حول 120 مطلق تزوجوا في نفس العام. وقد أوضحت هذه الدراسة أن متوسط طول فترة الزواج كان 8.5 سنة، بانحراف معياري 1.2 سنة. ما هو تقدير متوسط طول فترة الزواج لكل هؤلاء المطلقين مستخدماً مستوى ثقة 95%.

16- لمستوى الثقة التالية حدد ما يقابلها من درجة لِـ Z:

مستوى الثقة	ألفا α	مساحة أبعد من Z	درجة Z
% 95	0.05	0.0250	± 1.96
% 94			
% 92			
% 97			
% 98			
% 99.9			

الهوامش والمصادر:

أولا: الهوامش:

1- لمزيد من المعلومات: انظر:

EARL BABBie , The basic of Social Research , Wadsworth , USA , 2005 , PP. 212 - 213.

وعبد الله عامر الهمالي، أسلوب البحث الاجتماعي وتقنياته، منشورات جامعة قاريونس، بنغازي، 2003، ص ص 235 - 263.

2- EARL BABBie , op.cit , P.190.

3- George DiEkhoff, Statictics for the social and Behavioral Sciences: Univariate , Bivariate, Multivariate, MCGraw Hill, Companies , Inc. USA , 1992 , P. 86.

4- George Argyrous , Statistics for Social and Health Research , With a Guide to Spss , Sage Publications ,London , 2001 , PP.238 - 240.

5- Ibid , P. 240.

6- Ibid , PP. 248 - 252.

7- Ibid , PP. 252 - 255.

8- Ibid , PP. 255 - 257.

9- Ibid , PP. 259 - 260.

ثانياً: المصادر:

1. George Argyrous , Statistics for Social and Health Research , With a Guide to Spss , Sage Publications ,London , 2001.
2. George DiEkhoff, Statictics for the social and Behavioral Sciences: Univariate, Bivariate, Multivariate, MCGraw Hill, Companies , Inc. USA , 1992.
3. Frederick J. Gravetter and Larry B. Wallnau , Statistics for The Behavioral Sciences , 8eted , Wadsworth Cengate Learing , USA , 2010.
4. Joseph F. Healey, The Essentials of Statistitics: A Tool for Social Research, Wadsworuh Cengate Learning, USA. 2010.

الفصل الحادي عشر

اختبار الفروض: اختبار Z لمتوسط عينة واحدة

اختبار الفرض: فكرة عامة

في الفصل السابق، تناولنا التقنيات المتعلقة بتوزيعات المعاينة وتقدير معلمات المجتمع من إحصاء العينة. في هذا الفصل والفصول اللاحقة سوف نتناول بعض التطبيقات الأخرى المتعلقة بالإحصاء الاستدلالي ويطلق عليها اختبار الفروض أو دلالة الاختبار hyphothesis testing or signficance testing. في هذا الفصل سنتناول بالتفصيل التقنيات المتعلقة باختبار لمتوسط عينة واحدة. وكما رأينا في الفصول السابقة أن المعلومات المتحصل عليها من خلال العينة العشوائية ليست بالضرورة دائماً انعكاساً حقيقياً للمجتمع الذي سحبت منه هذه العينة.

ولتوضيح ذلك دعنا نسوق المثال التالي:

نفترض أن لدينا مجتمع مكون من 400 شخصٍ بمتوسط عمري يصل إلى 35 سنة وانحراف معياري يصل إلى 13 سنة. وقد تم سحب عينة من هذا المجتمع تصل إلى 150 شخصاً بمتوسط عمري 32 سنة ويود الباحث معرفة ما إذا كانت هذه العينة قد سُحِبَتْ من هذا المجتمع أم لا؟ إن الفرق العمري بين متوسط المجتمع ومتوسط العينة وصل إلى

ثلاث سنوات. هل هذا الفرق يمكن أن يقترح أن هذه العينة قد جاءت من مجتمع آخر أم أنها قد سحبت من هذا المجتمع بمتوسط عمري 35 سنة وأن الفرق في هذه السنوات الثلاث جاءت عن طريق التباين العشوائي عند المعاينة؛ بمعنى آخر، يمكننا القول، بأن هناك احتمالين لتفسير الطريقة التي اختلفت فيها نتائج العينة على نتائج المجتمع الذي تم فيه الاعتقاد بأن العينة لم يتم سحبها من هذا المجتمع.

فالتفسير الأول مفاده أن العينة قد تم سحبها فعلاً من المجتمع، ولكن هذه العينة تم سحبها بالصدفة، لعدد كبير من ذوي الأعمار الصغيرة. وفي هذه الحالة يمكننا أن نطلق على هذا التفسير لنتيجة العينة اسم الفرض الصفري Ho (لا يوجد فرق) ويمكن صياغة ذلك رياضياً:

$$Ho: u = 35 \text{ سنة}$$

أما التفسير البديل فمفاده أن العينة تم سحبها من مجتمع آخر متوسط أعمار أفراده ليست مساوية لـِ 35 سنة. ونطلق على هذا التفسير الفرض البديل Hi (يوجد فرق) وجبرياً يمكننا صياغته كالتالي:

$$Hi: u \neq 35 \text{ سنة}$$

إن هذين الفرضين مانعي التبادل، فإذا كان أحد هذين الفرضين صحيحاً فإن الآخر يكون غير صحيح، إما أن تكون العينة قد جاءت من مجتمع متوسط أعمار أفراده 35 سنة أو لا. وهذه تشبه حالة الملاكمة بين شخصين لابد لأحدهما أن ينتصر وبالتالي نحن لا نعرف أيّاً من الفرضين يكون صحيحاً.

السؤال الذي يمكن طرحه في إطار اختبار الفروض الإحصائية هو لِمَ نفترض أنه لا يوجد فرق لاعتقادنا أن هذا الافتراض الذي أوردناه افتراضاً صحيحاً. إن المنطق المتعلق باختبار الفروض يبين لنا ما إذا كانت فرضية " لا يوجد فرق " غير متساوقة مع نتائج البحث، وبالتالي فإن النقاش يقودنا إلى أن هذه الفرضية غير مبرَّرة. إننا نحاول أن نبرهن على أنه يوجد فرق وذلك من خلال إثبات العكس "فرضية لا يوجد فرق" ولفعل ذلك يحتاج الباحث إلى جملة من الإجراءات البحثية للوصول إلى هذه النتيجة، ومن خلال

الأسئلة التي نسوقها في متن هذا الكتاب؛ سوف تمكننا من اختبار هذه الافتراضات إذا كانت مقبولة أو متسقة مع البيانات البحثية. ومن خلال المثال الذي بين أيدينا فإنه – بشكل قوي – يساورنا الشك بأن العينة بمتوسط عمري 32 سنة لم تسحب من مجتمع بعدد 400 شخصٍ بمتوسط عمري 35 سنة. وبالرغم من الاعتقاد القوي أن هذا الفرض يكون صحيحاً، دعنا فعلياً نبدأ بالافتراض العكسي، وما نعتقده بأن هذا الفرض غير صحيح. وإذا ما بينا أنه من غير المحتمل لعينة بمتوسط عمري 32 سنة أن تسحب من مجتمع بمتوسط عمري 35 سنة، إذاً فإن هذه الفرضية التي بُدئي بها لا يمكن أن تكون مقبولة. وبالتالي يكون هناك مبرر لرفضها. ومن هنا جاءت عملية اختبار الفروض. نبدأ أولاً بصياغة الفرض الصفري Ho لاختباره بالمقارنة بالنتائج الفعلية لعينة الدراسة. وغالباً أننا لا ننجح في هذا الاختبار الذي نسعى إليه، دعنا نفترض لغرض النقاش، أن الفرض الصفري الذي مفاده "لا فرق" يعد فرضاً صحيحاً. ونحن نفترض أن العينة بمتوسط عمري 32، تم اختيارها من هذا المجتمع بمتوسط عمري 35 سنة بالرغم من وجود فرق ثلاث سنوات. والسؤال الذي يُطْرَح هنا هو: هل نتيجة هذه العينة ذات متوسط عمري 32 غير ثابتة مع فرضية أن المتوسط العمري للمجتمع يصل إلى 35 سنة؟ ما هي احتمالية الحصول على عينة يختلف متوسطاتها عن قيمة متوسط المجتمع 35 سنة بثلاث سنوات أو أكثر. هنا تدخل مسألة توزيع المعاينة لمتوسط العينات في مثل هذا الموقف. ونعني بتوزيع المعاينة توزيع المتوسطات للعينات الاحتمالية المعادلة والمتساوية في الحجم. كما يمكننا الإشارة إلى توزيع المعاينة والتي بإمكاننا معرفة خصائصها بالتفصيل ولتحديد احتمالية الحصول على عينة بمتوسط 32 سنة إذا كانت قيمة متوسط المجتمع 35 سنة. إن هذا الموقف يمدنا بأرجحية تسمح لنا أن نراهن إما على الفرض الصفري أو الفرض البديل. إن الأمر واضح لاستنتاج إجراء هذه الاحتماليات والذي أصبح الآن مألوفاً لدينا من خلال تحويل إحصاء العينة (المتوسط العمري) إلى درجة Z وبعد ذلك ننظر إلى احتمالية الارتباط من خلال النظر إلى جدول المساحة تحت المنحنى الطبيعي المعياري (جدول رقم 1).

جدول (11ـ1) المساحة تحت المنحنـى الطبيـعي المعياري

Table A1 Area under the standard normal curve

Z	Area under curve between both points	Area under curve beyond both points	Area under curve beyond one point
±0.1	0.080	0.920	0.4600
±0.2	0.159	0.841	0.4205
±0.3	0.236	0.764	0.3820
±0.4	0.311	0.689	0.3445
±0.5	0.383	0.617	0.3085
±0.6	0.451	0.549	0.2745
±0.7	0.516	0.484	0.2420
±0.8	0.576	0.424	0.2120
±0.9	0.632	0.368	0.1840
±1	0.683	0.317	0.1585
±2.1	0.964	0.036	0.0180
±2.2	0.972	0.028	0.0140
±2.3	0.979	0.021	0.0105
±2.33	0.980	0.020	0.0100
±2.4	0.984	0.016	0.0080
±2.5	0.988	0.012	0.0060
±2.58	0.990	0.010	0.0050
±2.6	0.991	0.009	0.0045
±2.7	0.993	0.007	0.0035
±2.8	0.995	0.005	0.0025
±2.9	0.996	0.004	0.0020
±3	0.997	0.003	0.0015

إن الخطوة الأولى حساب قيمة Z المرتبطة بنتيجة العينة. وعند حساب درجات Z للمثال المطروح، فإننا نستخدم المعادلة التالية:

$$Z = \frac{\overline{x} - u}{\frac{\sigma}{\sqrt{N}}}$$

وللعينة المسحوبة بـِ 150 شخصاً بمتوسط عمري 32 سنة، فإن درجة Z تكون:

$$Z = \frac{\bar{x} - u}{\frac{\sigma}{\sqrt{N}}} = \frac{32 - 35}{\frac{13}{\sqrt{150}}} = -2.8$$

إن هذه المعادلة في جوهرها قد بينت بشكل معياري فرق ثلاث سنوات بين درجة العينة والقيمة المفترضة للمجتمع وذلك بتحويلها إلى درجة Z. إن الميزة في إظهار الوحدات الطبيعية بذلك الفرق الذي تم قياسه مبدئياً (بالسنوات) يمكن أن نشير إليه الآن في جدول المساحة تحت المنحنى الطبيعي (انظر جدول 1) لنحدد احتمالية الحصول على درجة Z تساوي 2.8− أو أكثر.

قد يتعجب المرء لماذا الإشارة في هذا الصدد إلى العمود المتعلق بالمساحة تحت المنحنى الطبيعي أبعد من نقطتين، بدلاً من جدول المساحة تحت المنحنى الطبيعي أبعد من نقطة واحدة. إن هذا الأمر يتعلق بالكيفية التي صيغت بها المشكلة منذ البداية. فقد كان الاهتمام منذ البداية منصباً في معرفة احتمالية سحب عينة تختلف عن متوسط المجتمع الافتراضي بـِ 32 سنة أي بثلاث سنوات أو أكثر، حيث إن قيمة الفرق الذي تـحصلنا عليه من واقع العينة المسحوبة (بمتوسط عمري 32 سنة) قد سحبت من المجتمع بمتوسط عمري 35 سنة. وبما أن تباين المعاينة يمكن أن يتسبب في متوسطات العينات العشوائية إما إلى أعلى أو إلى أسفل مقارنة بمتوسط المجتمع؛ فمتوسط العينة يمكن أن يختلف كذلك بثلاث سنوات أو أكثر من القيمة المفترضة للمجتمع إما بثلاث سنوات أكثر (متوسط 38 سنة) أو بثلاث سنوات أقل (متوسط 32 سنة). وعليه فإننا نشير إلى عمود وسط الجدول لنقرر احتمالية السحب، من خلال تباين المعاينة فقط، إن المساحة تحت المنحنى أبعد من درجات ($Z = +2.8$) أو ($Z = -2.8$) تكون 0.005. هذه الاحتمالية للسحب من مجتمع بمتوسط عمري 35 سنة، وعينة بمتوسط عمري 3 سنوات فوق أو تحت هذا المتوسط. بمعنى آخر، فقط 5 مرات في 1.000 عينة سوف يختلف متوسط العينة عن متوسط المجتمع الذي يصل متوسطه إلى 35 سنة بثلاث سنوات أو أكثر. فالخيار لدينا حر، إما:

- البقاء على الفرضية التي مفادها أن العينة سحبت من مجتمع (بمتوسط عمري 35

سنة) بأن تكون صحيحة، وبالتالي تفسر نتائج العينة بشكل استثنائي (5 في كل 1000 حادثة) أو

- رفض الفرضية بأن العينة تم سحبها من مجتمع (متوسط عمري 35 سنة) وأن إحصاء العينة ليس استثنائياً، ولكنه بدلاً من ذلك يعكس أن العينة تم سحبها من مجتمع آخر يتضمن متوسط عمري أكثر من 35 سنة.

مفترضين أرجحية أكبر من الخيار الأول هو الصحيح، وعليه يمكننا المراهنة على رفض الفرض الذي مفاده أن العينة قد سحبت من مجتمع بمتوسط عمري 35 سنة. إن الفرق بين نتائج القيمة الافتراضية للمجتمع كبيرة (ثلاث سنوات) إلى درجة يكون معها من غير المحتمل أنها جاءت عن طريق التباين العشوائي عند المعاينة، غير أنها بدلاً من ذلك تعكس أننا لم نقم بعملية المعاينة من مجتمع بمتوسط عمري 35 سنة. ولتوضيح هذا الإجراء مرة ثانية دعنا نفترض أن عينة من 150 شخصاً تعطينا النتيجة التالية:

$$\overline{X} = 36 \text{ سنة}$$

مرة ثانية، يمكننا أن نفترض أن العينة تم سحبها من مجتمع بمتوسط عمري 35 سنة. وبشكل واضح فإن الفرق مرة ثانية بين إحصاء العينة ومعلمة المجتمع، يصل هذه المرة إلى سنة. (1=35−36). إن التناقض هنا يبدو واضحاً بين فرضيتنا القائلة بأن العينة تم سحبها من مجتمع بمتوسط عمري 35 سنة وبين مشاهدتنا بأن نتيجة العينة ليست مساوية تماماً للقيمة الإحصائية المتعلقة بالمجتمع. هل هذه النتيجة تقودنا إلى رفض الفرضية بدلاً من القول بأن العينة تم سحبها من مجتمع آخر؟

للإجابة على هذا التساؤل نحتاج إلى استنتاج احتمالية الاختبار العشوائي لعينة تختلف متوسطاتها عن متوسط مجتمع يقدر بـِ 35 سنة، يزيد بسنة أو أكثر. إننا نحتاج أولاً إلى تحويل نتيجة العينة إلى درجة Z:

$$Z = \frac{\overline{x} - u}{\frac{\sigma}{\sqrt{N}}} = \frac{36 - 35}{\frac{13}{\sqrt{150}}} = 0.9$$

ويشير جدول المساحة تحت المنحنى إلى أن احتمالية الحصول على هذه الدرجة لـِ Z أو أكبر على أي من جانب المتوسط هي 0.368 (انظر جدول 1).

مثال: نفترض أن متوسط المجتمع وصل إلى 35 سنة، وأن متوسط العينة وصل إلى 36 سنة، يعني هذا أن المجتمع بمتوسط عمري 35 سنة قريباً من 37 / 1000 عينة سيكون متوسطها العمري مختلفاً عن 35 سنة أو أكثر. إن التباين العشوائي سيسبب تقريباً ⅓ كل العينات لتتباين إلى حدٍ كبير من المجتمع بقيمة متوسط 35 سنة. مثل هذه الاحتمالية العالية، تقودنا إلى القول بأن نتيجة العينة هي ببساطة راجعة إلى تباين عشوائي عندما تم اختبار هذه العينة من مجتمع (بمتوسط عمري 35 سنة) [1].

نموذج الخطوات الخمس لاختبار الفرض:

يتطلب التعامل مع أي مشكلة متعلقة بالاستنتاج الإحصائي أن تتبع نموذج الخطوات الخمس لاختبار الفروض الإحصائية. وهذه الخطوات:

1- صياغة الفرض الصفري والفرض البديل.
2- اختيار درجة الدلالة الإحصائية.
3- حساب إحصائي الاختبار.
4- بيان المنطقة الحرجة أو منطقة الرفض.
5- اتخاذ القرار.

في متن هذا الفصل سنبين هذه الخطوات بشكل مفصل قبل أن نعطي أمثلة لتوضيحها:

الخطوة الأولى: صياغة الفرض الصفري والفرض البديل:

عادة ما يبدأ الباحث بسؤال بحثي، يحاول الإجابة عليه على أن تكون هذه البداية واضحة ودقيقة، ومصاغة في شكل سؤال يقود إلى إجراء البحث ويضمن الحصول على المعرفة الحقيقية. إذاً تكمن النقطة الجوهرية منذ البداية في صياغة المشكلة البحثية صياغة واضحة حول القضية المطروحة قبل أي تحليلات إحصائية. وتساعد هذه العملية الباحث في تحاشي كم هائل من البيانات التي قد تكون غير ضرورية للسؤال الذي تم طرحه.

فعلى الباحث منذ البداية أن يكون واضحاً في تحديد أهداف بحثه ليستطيع في مرحلة لاحقة استخدام التحليلات الإحصائية الملائمة. وعلى النقيض من ذلك، إذا ما عجز الباحث عن تحديد أهداف بحثه منذ البداية بشكل دقيق، فإن ذلك يضعه في موقفٍ يكون فيه غير قادر على تحديد الإجراءات الإحصائية الملائمة لتحليل بياناته التي تم جمعها.

والأمثلة اللاحقة تبين لنا كيف يمكن لباحث أن تكون أسئلته البحثية التي يطرحها واضحة ودقيقة:

- هل الجماعات الأصغر سناً يختلفون عن الجماعات الأكبر سناً في نظرتهم حول تحرر المرأة؟
- هل خريجو الثانوية الاجتماعية يختلفون عن خريجي الثانويات العامة في معدلات التحصيل الدراسي الجامعي؟
- هل أطفال المؤسسات الإيوائية يختلفون عن أطفال الأسر الطبيعية في التوافق الاجتماعي والشخصي؟

يتضح لنا من خلال هذه الأسئلة المطروحة أنها تحتوي على عناصر مهمة فهي توضح:

- المجتمع أو "المجتمعات" التي تريد أن نضع صياغة حولها.
- المتغيرات التي يود الباحث أن يجمع البيانات حولها.
- الإحصاءات الوصفية المناسبة لوصف البيانات.

ففي السؤال الأول من الأسئلة المطروحة على سبيل المثال، يتضح لنا بما لا يدع مجالاً للشك، أننا نرغب في الحصول على بيانات من مجتمعين. الجماعات الصغيرة في السن والجماعات الأكبر سناً. أما في السؤال الثاني، فالباحث يرغب في معرفة الفرق في معدلات التحصيل الجامعي. وأخيراً، أطفال المؤسسات في مقابل أطفال الأسر الطبيعية.

أـ الفرض الصفري الذي مفاده لا فرق Ho:

إن هذه الصياغة تعني أن الإحصاء المستخدم لوصف المجتمع الخاضع للدراسة ستكون

قيمته متساوية ومحددة سلفاً؛ على أن يكون الفرض الصفري قابلاً للرفض أو الإثبات؛ وهذا يعني أنه من الممكن أن يكون الفرض خاطئاً، بحيث لا يكون هناك أي غموض: إما أن يكون إحصاء المجتمع الذي نتعامل معه يمتلك قيمة محددة، في إطار التعريفات الإجرائية للمتغير أو أنه لا يمتلك مثل هذه القيم. إن صياغة الفرض الصفري جبرياً يعتمد على الإحصاء الوصفي المستخدم. فإذا كان الباحث يتعامل مع فرضية حول متوسط المجتمع، فإن المعادلة المستخدمة على سبيل المثال، تكون:

$$Ho: u = X$$

حيث إن X هي قيمة الاختبار المحدد سلفاً. ففي المثال السابق، كان الاختبار يتعلق بما إذا كانت قيمة u تساوي 35 سنة. إن السؤال الذي يمكن طرحه هو: من أين جاءت صياغة قيمة الاختبار في الفرض الصفري؟ للإجابة على هذا السؤال يتطلب منا النظر إلى نوعين مختلفين من الأسئلة التي تقودنا لاستقصاء ما إذا كانت معلمة المجتمع تأخذ قيمة محددة.

السؤال الأول: هل اختيار القيمة المبينة تم بناؤها على أسباب عملية أو سياسية. والسؤال الثاني: هل يرغب الباحث في مقارنة المجتمع تحت الدراسة مع مجتمع آخر تكون فيه قيمة المعلمة معلومة. فالباحث على سبيل المثال، يرغب في المقارنة بين مجتمع A ومجتمع B، فيما يتعلق بمتوسط كمية مشاهدة الإذاعة المرئية في اليوم الواحد: فالأطفال في مجتمع A بين 5 و 12 سنة والأطفال في مجتمع B تتراوح أعمارهم بين 5 و 12 سنة. فالباحث قد يكون على دراية بالإحصاءات في كلا المجتمعين.

ولما كان متوسط مشاهدة الإذاعة المرئية لدى أطفال مجتمع A غير معلوم فإن الباحث يود أن يستنتج ما إذا كان متوسط هؤلاء الأطفال مساوياً لمتوسط أطفال B المعلوم.

ب- الفرض البديل Hi:

مفاده أن معلمة المجتمع ليست مساوية للقيمة التي تم تحديدها سلفاً أي يوجد فرق $Hi: u \neq x$. وتجدر الإشارة، إلى أن الفرض البديل يمكن أن يتخذ صياغة أخرى دقيقة

فبدلاً من القول بأنه يوجد فرق، فإن الباحث بإمكانه الاعتقاد في أن هذا الفرق سيتخذ اتجاهاً محدداً. على سبيل المثال، قد يشك في أن متوسط مشاهدة للإذاعة المرئية في مجتمع A لا يعني أنه يختلف عن متوسط كمية مشاهدة الإذاعة المرئية لدى أطفال مجتمع B ولكنهم يشاهدون أقل. وخلافاً لذلك يمكننا القول، بأن متوسط ما يشاهدونه أكثر في اليوم في كلتا الحالتين ليس الأمر متعلقاً بالفرق بين متوسط المجتمعين، وإنما هناك اتجاهية في هذا الفرق [2]، ويمكننا جبرياً صياغة ذلك كالتالي:

Hi: < X

أو Hi: > X

الخطوة الثانية: اختيار اختبار الدلالة:

تتوفر لدينا مجموعة من الاختبارات التي تساعدنا في اتخاذ القرار بين الفرض الصفري والفرض البديل. وأن كل هذه الاختبارات تتطلب عينات عشوائية ولكنها تتباين طبقاً للمعلومات المتوفرة لدى الباحث.

في هذا الفصل سنقدم الاختبار الأساسي المتعلق باختبار Z لمتوسط عينة واحدة. عادة ما يطلق على دلالة الاختبار هذه، اسم مختصر يستند على الإحصائي الذي ابتكر هذه الاختبارات كاختبار ولكوكسن Wilcoxon Test إجمالاً، توجد مجموعة من العوامل المهمة تحدد اختيارنا لاختبار الدلالة وهذه العوامل:

- عدد المجتمعات و"العينات" التي من خلالها يتم الاستدلال.
- الإحصاء الوصفي المستخدم لوصف البيانات الخام، التي تعتمد على مستوى القياس؟
- فيما إذا كان لدينا متغيرات مستقلة أو متغيرات تابعة.

إن كل اختبار بكلمات أخرى، يعتمد في تطبيقاته على ظروف محددة. انظر الجداول التالية التي تبين لنا بشكل سريع دليلاً لاختيار اختبار الدلالة الملائم؛ اعتماداً على هذه العوامل الأساسية. وتجدر ملاحظة أن هذه الإجراءات المتعلقة باختبار الدلالة لا تشتمل على كل الاختبارات الممكنة والمتوفرة، وإنما تقدم تلك الاختبارات التي تم التعامل معها في هذا الكتاب.

جدول (11ـ2) اختبارات الدلالة: حالة العينة الواحدة

مستوى القياس	اختبار الدلالة
الاسمي / الترتيبي	– اختبار Z للنسب الثنائية. – اختبار X^2 اختبار حسن المطابقة للتوزيع التكراري.
ذي المسافات والنسبي	– اختبار Z لمتوسط (تباين مجتمع معروف). – اختبار t لمتوسط (تباين مجتمع غير معروف).

جدول (11ـ3) اختبار الدلالة: لعينتين مستقلتين

مستوى القياس	اختبار الدلالة
الاسمي	– اختبار X^2 للاستقلال لجداول التوافق (يمكن أيضاً استخدام اختبار Z لنسب التوزيع الثنائي).
الترتيبي	– اختبار Z لمجموع الرتب (والذي يعرف أيضاً باختبار Willcoxon test والذي هو مرادف لاختبار مان–وتني u.
ذو المسافات والنسبي	– اختبار t لمتوسطين متساويين.

جدول (13ـ4) اختبار الدلالة: لأكثر من عينتين مستقلتين

مستوى القياس	اختبار الدلالة
الاسمي	– اختبار X^2 للاستقلال
الترتيبي	– اختبار كروسكل وليز H
ذو المسافات والنسبي	– أنوفا Anova: اختبار F لمتوسطين متساويين

جدول (13ـ5) اختبار الدلالة: لعينتين غير مستقلتين

مستوى القياس	اختبار الدلالة
الترتيبي	– اختبار ولكوسكن
ذي المسافات والنسبي	– اختبار t لفرق المتوسط

المصدر: George Argyrous, op.cit , p. 271

آخذين في الاعتبار أنه عند استخدام هذه الجداول، يتوجب على الباحث أن يأخذ في اعتباره أن أي اختبار يمكن تطبيقه على مستوى محدد من القياس، كما يمكنه أيضاً تطبيقه على مستوى أعلى من القياس. فالاختبارات الموضحة في جدول (2) والمتعلقة بالبيانات الاسمية يمكن تطبيقها على بيانات المستوى الترتيبي وذي المسافات والنسبي. كذلك الحال في الاختبارات المتعلقة بالمستوى الترتيبي يمكن تطبيقها على بيانات مستوى ذي المسافات والنسبي.

في هذا الفصل - كما أشرنا - سنتناول اختبار Z لمتوسط عينة واحدة قبل تطبيق هذا الاختبار، تجدر الإشارة إلى الأحوال التي تسمح لنا بتطبيق هذا الاختبار:

لابد أن يكون هناك سؤال بحثي نرغب من خلاله في معرفة النـزعة المركزية لمتغير ما، ومستوى الدلالة (ذو المسافات/النسبي) والإحصاءات الوصفية المناسبة لتلخيص بيانات العينة (المتوسط الحسابي).

على طول المتغير لابد أن يكون المجتمع موزعاً توزيعاً طبيعياً و / أو حجم العينة يجب أن يكون كبيراً (N > 100). إن أحد هذين الشرطين الأخيرين طبقاً للنظرية الأساسية في تقارب التوزيعات الإحصائية سيؤمن لنا أن تكون توزيعات المعاينة لمتوسطات العينة توزيعات طبيعية.

الخطوة الثالثة: حساب درجة العينة:

نعني بحساب درجة العينة العملية التي يتم بها تحويل الإحصاء الوصفي المناسب إلى درجة معيارية (Standardized Score) مثل درجة Z وذلك من خلال تعويض بيانات العينة في المعادلة المناسبة. وكما بينا أعلاه، يمكننا تحويل متوسط عينة إلى درجة Z من خلال المعادلة التالية:

$$Z_{\text{العينة}} = \frac{\bar{x} - u}{\sigma / \sqrt{N}}$$

ومن خلال جدول المساحة تحت المنحنى الطبيعي يمكننا حينئذٍ أن نقرر احتمالية الحصول على درجة خاصة بـِ Z إذا كان الفرض الصفري في الحقيقة فرضاً صحيحاً. ونعني بهذا دلالة إحصاء العينة، التي يطلق عليها بشكل واسع P.Value.

الخطوة الرابعة: بيان الدرجة الحرجة أو المنطقة الحرجة:

عند أي نقطة تصبح قيمة العينة كبيرة جداً، أو ارتباطها مع القيمة الاحتمالية يصبح صغيراً جداً، تمكننا من رفض الفرض الصفري، ويمكننا تقرير ذلك بتحديد نقطة قاطعة (a cut off point) ترسم الدرجات العليا والدرجات الدنيا.

إن هذه الخطوة تنقسم إلى ثلاثة أجزاء:

أ- اختبار مستوى الدلالة: وهي القيمة المحددة التي على ضوئها يُقْبَلُ أو يُرْفَضُ الفرض الصفري. فإذا استخدمنا التحليل السابق المتعلق بالمتوسط العمري لعينة من المجتمع، فالقرار ما إذا كنا نقبل أو نرفض الفرض الصفري كان قراراً سهلاً، ففي المثال الأول، الذي يحتوي على متوسط عمري 32 سنة، فالاحتمالية بأن هذه العينة قد سحبت من مجتمع بمتوسط عمري قدره 35 سنة كانت صغيرة. بينما في حالة المثال الثاني الذي يصل في متوسط العينة إلى 36 سنة فالاحتمالية كانت كبيرة جداً. ولكن السؤال الذي يمكن طرحه هنا هو: ما النتيجة إذا ما وقعت العينة في مكان ما بين الاثنين؟ عند أي نقطة ينتج عنها الحصول على احتمالية صغيرة إلى حدٍ كافٍ يمكننا من خلالها القول بأن الفرض الصفري ليس صحيحاً؟ وبتحديدنا للنقطة القاطعة التي يطلق عليها اختيار مستوى الدلالة، أو مستوى ألفا α التي يشار إليها بـِ 0.05 مستوى دلالة: $\alpha = 0.05$.

ولفهم هذه القضايا المتعلقة باختيار مستوى الدلالة يجب التمييز بين نوعين من الخطأ هما: الخطأ من النوع الأول Type I Error والخطأ من النوع الثاني Type II Error (خطأ بيتا Beta Error)[3].

الخطأ من النوع الأول: يحدث الخطأ من النوع الأول عندما يتم رفض الفرض الصفري لا فرق، وبالرغم من الحقيقة أنه لا يوجد فرق. ولتقييم ما إذا كانت العينة في المثال السابق بمتوسط عمري 32 سنة وقد تم سحبها من مجتمع بمتوسط عمري 35 سنة، فإننا قد رفضنا الفرض الصفري الذي مفاده لا يوجد فرق. إن الفرص لاختيار من مجتمع بمتوسط عمري 35 سنة، وعينة بمتوسط عمري 32 سنة أو أقل هي فقط 25 في 10.000. بمعنى آخر، إذا أخذنا 10.000 عينة من مجتمع متوسط أعماره 35 سنة، فإننا

ستحصل فقط على 25 سيكون متوسط أعمارهم 32 سنة أو أقل. وعلى أية حال، قد نكون في حقيقة الأمر قد اخترنا هؤلاء عن طريق الحظ 25 في 10.000، فالعينة قد تكون حقاً سحبت من مجتمع بمتوسط عمري 35 سنة، لكن العينة تماماً قد تم سحبها بشكل عشوائي لحالات قليلة من ذوي الأعمار الصغيرة. (توجد دائماً هذه الصدفة في مثل هذه الحادثة). هذا هو الأمر الذي يجعلنا نتحدث عن الاحتماليات. فهو سؤال متعلق باحتمالية الصدف التي نكون بها مهيئين للوقوع في هذا الخطأ.

الخطأ من النوع الثاني: يحدث الخطأ من النوع الثاني عندما يتم قبول الفرض الصفري وهو في الحقيقة غير صحيح. فعلى سبيل المثال، عندما كان المتوسط العمري للعينة 36 سنة، فقد توصلنا إلى قرار مفاده أن هذه العينة لم يتم سحبها من مجتمع بمتوسط عمري 35 سنة. إن الفرق بين إحصاء العينة وقيمة المعلمة المفترضة يكون صغيراً جداً حيث يمكننا إرجاعه إلى التباين العشوائي. وعلى أية حال، فإنه يمكن في الواقع أن يكون المجتمع الذي سحبت منه العينة لم يصل متوسط أعمار أفراده 35 سنة، ولكن العينة تماماً قد حدث أن تم اختيار أفرادها بطريقة غير ممثلة.

إن العلاقة بين هذين الاحتمالين من الأخطاء يمكن تلخيصها في الجدول التالي:

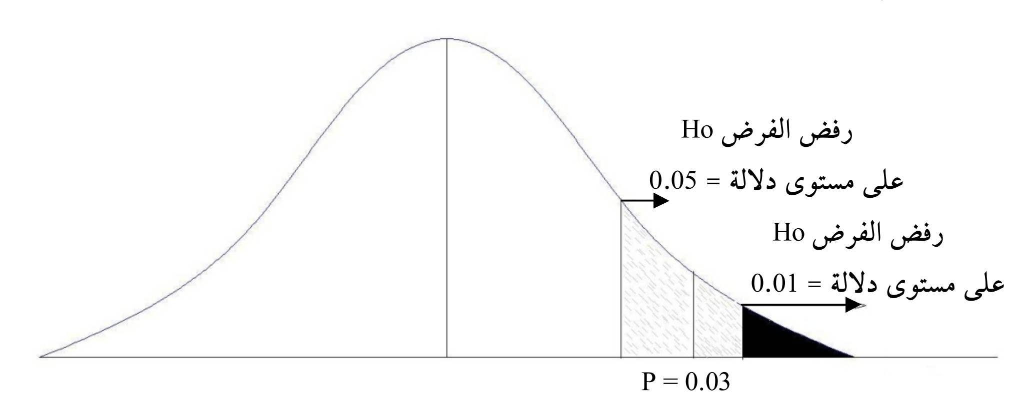
جدول رقم (11 ـ 6) المخرجات الممكنة للقرار الإحصائي

	موقف فعلي		
	لا تأثير Ho صحيح		وجود تأثير H0 خطأ
القرار	رفض الفرض Ho	الخطأ من النوع I	قرار صحيح
	قبول الفرض Ho	قرار صحيح	الخطأ من النوع II

من الواضح هنا أنه يوجد لدينا نوعان من الأخطاء مضادان لبعضهما البعض إلى حد أنه بتقليل الخطأ في أحدهما يؤدي إلى زيادة احتمالية الخطأ في الآخر. إن السؤال في واقع الأمر يتعلق بالخطأ الذي يتوجب علينا تحاشيه، وهذا في واقع الأمر يعتمد على

السؤال البحثي المطروح. فإذا كان الباحث يرغب في اختبار دواء جديد ومعرفة المضاعفات الجانبية لهذا الدواء، فالباحث هنا لابد أن يكون متأكداً من أن هذا الدواء ذا فعالية. بمعنى آخر، إن الباحث لا يريد أن يقع في الخطأ من النوع الأول (مبيناً أن الدواء يعمل فرقاً في حين أنه لا يعمل فرق) بسبب النتائج التي قد تكون مأساوية. إن الفرق في معدل التحسن المشاهد بين المجموعة التجريبية التي تناولت الدواء والمجموعة الضابطة التي ليس لديها النية أن تكون كبيرة جداً قبل الحديث بأن هناك تحسناً كبيراً لا يرجع إلى الصدفة (ولنقل 1 في 1000). إن ما نفعله هنا أننا نختار درجة دلالة 0.001 قبل رفض الفرض الذي مفاده أن الدواء لا يعمل فرقاً [4].

وتجدر الإشارة إلى أنه يمكننا أن نلاحظ الفرق إذا اخترنا درجة دلالة مختلفة. (انظر الشكل رقم 1).



شكل (11ـ1) مناطق الرفض على مستوى 0.05 و 0.01
(اختبار أحادي الجانب)

إن نتيجة العينة لديها احتمالية ظهور بـِ 0.03، إذا كان الفرض الصفري (لا فرق) فرضاً صحيحاً. فإنه يبدو واضحاً أن مستوى الدلالة الذي تم اختياره سوف يحدد لنا ما إذا كنا نرفض أو نقبل الفرضية التي مفادها (لا يوجد فرق). فإذا اخترنا على سبيل المثال، درجة دلالة $\alpha = 0.01$ حينئذٍ فإن الفرق بين قيمة العينة وقيمة المجتمع يمكن إرجاعها إلى التباين العشوائي: أي أننا في هذه الحالة لا نستطيع رفض الفرض الصفري.

ولكن إذا اخترنا درجة دلالة 0.05 عندئذٍ فإن نفس الفرق بين قيمة العينة وقيمة المعلمة المفترضة ستقودنا إلى رفض الفرض الصفري.

اختبار أحادي الجانب أو ثنائي الجانب:

إن اختيار اختبار أحادي الجانب أو ثنائي الجانب يتوقف في الأساس على توقعات الباحث حول المجتمع التي سحبت منه العينة. وهذه التوقعات تنعكس في فرضية البحث Hi التي تناقض الفرضية الصفرية Ho. وبالنظر إلى الخطوة الأولى من إجراءات اختبار الفروض المتعلقة بصياغة الفرض البديل الذي مفاده أن العينة تم سحبها من مجتمع قيمة متوسطه تختلف عن القيمة المفترضة وأن العينة تم سحبها من مجتمع قيمة متوسطه أكبر أو أصغر من القيمة المفترضة.

إن السؤال المطروح هنا هو: هل اهتمامنا ينصبُّ على الفرق أم أنه يتعدى ذلك إلى اتجاه هذا الفرق كذلك. في العينة التي طرحناها على سبيل المثال، فإننا نشك في الفرضية البديلة في مقابل الفرضية الصفرية التي مفادها لا يوجد فرق في أن العينة جاءت من مجتمع متوسط أعمار أفراده أصغر من 35 سنة. وعليه فإننا في هذه الحالة نجري اختبار أحادي الجانب لأننا نرغب في معرفة ما إذا كانت نتيجة العينة تقع في مكان بعيد إلى حد كافٍ إلى اليسار من متوسط المجتمع.

دعنا نقول إنه لا يوجد لدينا أي مبرر مسبق بالاعتقاد أن هذه هي الحالة: إن العينة يمكن أنها جاءت من مجتمع إما أن يكون المتوسط العمري لأفراده أصغر أو أكبر. في هذه الحالة فإن المنطقة الحرجة تنقسم إلى نهايتين –كل واحدة من هاتين النهايتين عند نهاية توزيع المعاينة – تلك الطريقة الفعالة التي بيناها في المثال السابق.

إنه عندما نحدد اتجاه الفرق في الفرض البديل حينئذٍ يتوجب علينا استخدام اختبار أحادي الجانب، ويجدر بنا أن ننوه بشكل دقيق إلى الجانب الملائم لتوزيع المعاينة. فإذا كان الفرض البديل أبقى بأن قيمة المجتمع سوف تكون أقل من القيمة المحددة، فالمنطقة الحرجة ستكون في الجانب الأيسر؛ أما إذا بقيت قيمة المجتمع أنها ستكون أكبر من القيمة المحددة، عندئذٍ فإن النقطة الحرجة ستكون في الجانب الأيمن. انظر الجدول (11 – 7).

جدول (11 ـ 7) اختيار اختبار أحادي الجانب

الفرض البديل	ديل توزيع المعاينة
$Hi : u \neq X$	كلا الجانبين (الديلين)
$Hi : u < X$	الجانب الأيسر من التوزيع
$Hi : u > X$	الجانب الأيمن من التوزيع

إن اختبارات الجانب الأيسر: عادة ما تستخدم عندما يرغب الباحث في تحقق ما إذا كان بعض متطلبات الحد الأدنى قد تم الإيفاء بها، في المقابل فإن اختبار الجانب الأيمن يستخدم عندما يرغب الباحث في معرفة ما إذا كانت بعض الحدود العليا أو المعيار لم يتجاوز هذه المتطلبات، على سبيل المثال، إذا أراد الباحث معرفة ما إذا كان متوسط الحياة لبعض الأدوات المستخدمة في المستشفى قد تم استخدامها على الأقل 4.5 سنة، فإننا في هذه الحالة نستخدم الجانب الأيسر من الاختبار. أما إذا كنا في نرغب ما إذا كان الوقت الذي تم فيه أخذ الدواء كان له تأثير على المريض في وقت لا يزيد عن 1.5 دقيقة، عندئذٍ يمكننا أن نستخدم الجانب الأيمن من الاختبار.

اشتقاق أو تحديد الدرجة أو الدرجات الحرجة:

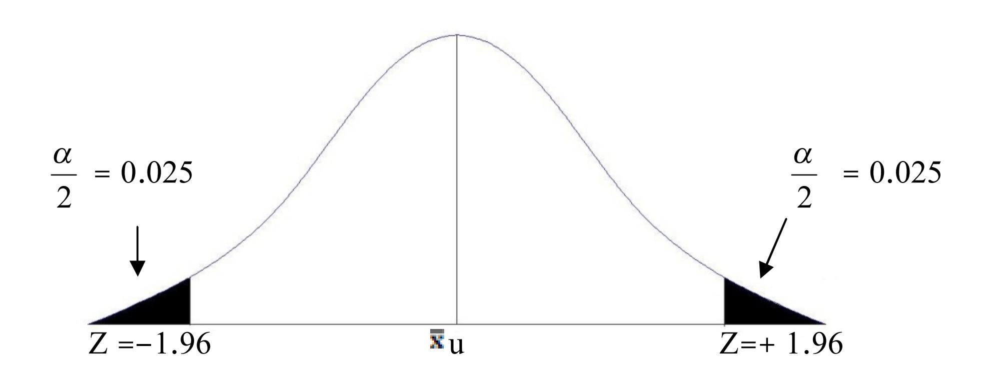
عندما يتم تحديد مستوى الدلالة، ونقرر ما إذا كنا نريد أن نستخدم اختبار أحادي الجانب أو ثنائي الجانب، حينئذٍ يمكننا اشتقاق درجات الدلالة (أحياناً نطلق عليها إحصاء الاختبار) التي تشير إلى منطقة الرفض. ونعني بمنطقة الرفض أو المنطقة الحرجة: مدى الدرجات التي على ضوئها يتم رفض الفرض الصفري. ولتحديد هذه الدرجات الحرجة يمكننا الرجوع إلى جدول المساحة تحت المنحنى الطبيعي المعياري:

$$\alpha \rightarrow Z \text{ الحرجة}$$

فعند اختيارنا لاختبار ثنائي الجانب يمكننا الإشارة إلى العمود في المساحة تحت

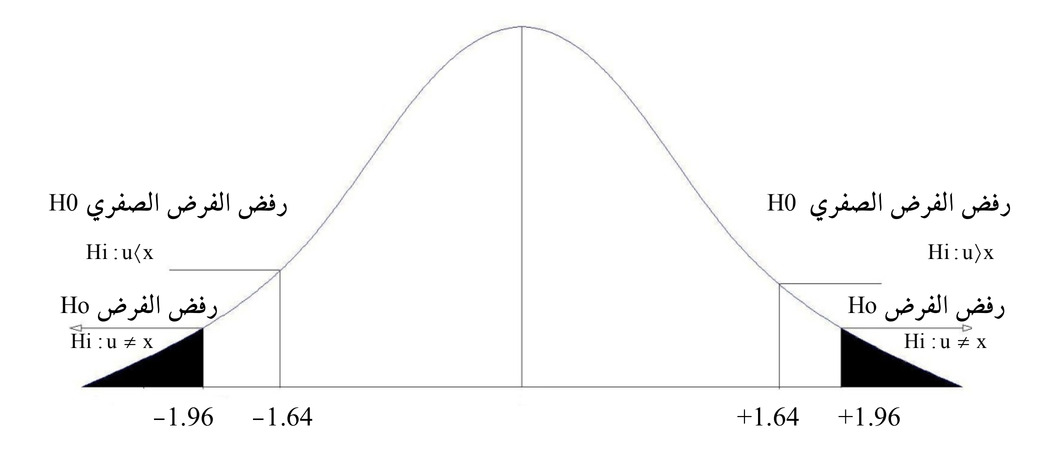
المنحنى أبعد من كلتا النقطتين، ونقرأ القيمة الموجبة والقيمة السالبة لدرجة Z (انظر الجدول رقم 1)؛ في حين يمكننا الإشارة إلى عمود المساحة تحت المنحنى أبعد من نقطة واحدة عند التعامل مع الاختبار أحادي الجانب، ونقرأ القيمة الموجبة والقيمة السالبة لدرجة Z استناداً على الجانب الملائم. على سبيل المثال، إذا كانت درجة Z تشير إلى المنطقة الحرجة بدرجة دلالة 0.05 α على اختبار ثنائي الجانب. فإنها تصل إلى 1.96 ±.

إن احتمالية 0.05 تُقَسَّم إلى منطقتين متساويتين في الحجم، كل واحدة في أي من جانبي التوزيع. أبعد درجة لـِ Z 1.96+ هي 0.025 أو 25 % للمساحة تحت المنحنى، وأن أبعد درجة لـِ Z 1.96− تقع ثانية في 0.025 المساحة تحت المنحنى (انظر شكل 11 – 2).



شكل (11 ـ 2) المناطق الحرجة لاختبار ثنائي الجانب
على مستوى دلالة 0.05 = α

إن الفرق بين الاختبار أحادي الجانب واختبار ثنائي الجانب يمكن توضيحه في الشكل التالي [3].



شكل (11 ـ 3) المناطق الحرجة لاختبار أحادي الجانب وثنائي الجانب على مستوى $\alpha = 0.05$

ففي اختبار أحادي الجانب لمستوى دلالة 0.05، المنطقة الحرجة تحت توزيع المعاينة تبدأ إما 1.645− أو 1.645+ ولكن ليس الاثنين اعتماداً على اتجاه الفرق الذي يعبر عنه الفرض البديل. أما في حالة اختبار ثنائي الجانب فإن منطقة الرفض تنقسم إلى قسمين أو نصفين لأن الباحث يرغب في نتيجة عينة إما أن تكون هذه النتيجة كبيرة جداً أو صغيرة جداً إذا ما قورنت بقيمة المجتمع. هذه العملية تحرك درجة Z نحو الخارج إلى 1.96 ±.

والذي نود الإشارة إليه أن هناك مستويات للدلالة مألوفة ومتفق عليه في سياق البحث، والمرتبطة بدرجة Z الحرجة لهذه المستويات من الدلالة.

وتصبح هذه المستويات من الدلالة أكثر ألفة عند إجراء البحوث بشكل مستمر. فإذا مارس الباحث بشكل كافٍ الإحصاءات الاستدلالية، فإنه بالتالي يـحتاج إلى معرفة هذه الدلالات وفي المحصلة النهائية يتوجب عليه حفظها، لاسيما مستوى الدلالة 0.05 والتي تستخدم بشكل شائع في مجال البحث الاجتماعي.

جدول (11ـ8) إيجاد الدرجة الحرجة الشائعة لدرجات Z
لاختبار أحادي الجانب (لمتوسطات عينة مفردة)

ألفا α	قيم الذيلين	قيم ذيل واحد	
		الذيل الأعلى	الذيل الأدنى
0.10	±1.65	+1.29	−1.29
0.05	±1.96	+1.65	−1.65
0.01	±2.58	+2.33	−2.33
0.001	±3.29	+3.10	−3.10

الخطوة الخامسة: اتخاذ القرار: مقارنة درجة العينة بالدرجة الحرجة:

بعد إتمام كل الإجراءات السابقة، حينئذٍ يمكننا أن نقرر إما أن نرفض الفرض الصفري أو لا نرفضه. لاحظ أن النتيجة تكون دائماً مصاغة في ضوء الفرض الصفري: نرفض أو لا نرفض. وعند اتخاذ القرار فإننا نقوم بواحدة من مقارنتين:

مقارنة Z العينة بـِ Z الحرجة.

أو مقارنة القيمة الاحتمالية p بمستوى ألفا α

إن الطريقة التي نصل بها إلى اتخاذ القرار يمكن تلخيصها في الجدول التالي:

جدول (11-9) ملخص إجراء اختبار الفروض

درجات العينة خطوة (3)	الدرجات الحرجة خطوة (4)	اتخاذ القرار خطوة (5)
$\bar{x}$ ↓		
Z العينة	Z الحرجة	إذا كانت Z العينة أبعد من 0 بالمقارنة بـ Z الاحتمالية نرفض الفرض الصفري
↓	↑	
جدول المساحة تحت المنحنى الطبيعي		أو
↓	↑	
P	α	إذا كانت P أصغر من α نرفض الفرض Ho

قد يكون من المفيد عند مرحلة اتخاذ القرار أن نقوم بمخطط المنحنى الطبيعي ونبين الدرجات الحرجة ودرجات العينة (كما بينا أعلاه). وهذه العملية تعطينا مؤشراً سريعاً ما إذا كنا نرفض الفرض الصفري؛ إذا كانت درجة العينة قد وقعت في منطقة الرفض (إذا وقعت القيمة الإحصائية في منطقة الرفض، حينئذٍ نرفض الفرض الصفري والعكس بالعكس).

وكثيراً ما يحدث إرباك عند نقطة اتخاذ القرار. وعند ملاحظة القيمة الاحتمالية لدرجة العينة، فالباحث عادة ما يصيبه الفزع إذا كانت القيمة الاحتمالية قريبة جداً من الصفر، باعتبار أن الباحث يظن أن الأعداد الصغيرة لا تشير إلى شيء، وعليه فإن الفرق الذي يطمح إليه أو لإيجاده قد يتبدد. ولكن الأمر هنا يكون صحيحاً. في العادة نريد أن نجد قيمة احتمالية منخفضة (منخفضة أكثر من مستوى ألفا)، وبما أن هذه القيمة

المنخفضة تشير إلى الفرض الصفري (لا فرق) ينبغي أن يرفض. وعلى الجانب الآخر، فالدرجة العالية للقيمة الاحتمالية تشير إلى أن الفرض الصفري (لا فرق) يجب ألا يرفض.

وعند إعداد، تقرير حول النتائج المتعلقة باختبار الفرض، حتى وإن كنا على ثقة بالقرار الذي نتخذه حول الفرض الصفري، فالمعلومات الكافية يجب أن تكون متوفرة حتى يستطيع القارئ أن يصدر حكمه الخاص. وعلى وجه الخصوص، فالاحتمالية الصحيحة مرتبطة بدرجة Z للعينة. يجب أن توصف، ليست بمجرد ما إذا كانت أعلى أو أقل من مستوى ألفا الذي تم اختيارها. هذا الأمر يسمح للقراء أن يقرروا ما إذا كانوا يشعرون بأنه الاحتمالية صغيرة بشكل كافٍ لتعطي المسوغ لرفض الفرض الصفري. إن اتخاذ القرار يترك للقارئ، بدلاً من الشخص الذي أجرى الاختبار: فالقارئ يمكنه مقارنة الاحتمالية بمستوى ألفا التي يعتقد أن تكون مسوغاً بسياق البحث أكثر من إخباره ببساطة أن النتيجة دالة على مستوى 0.05، وإذا ما استمرينا في بيان ما تم تقريره، فإن احتمالية العينة يمكن أن تكون 0.049 أو 0.00001. إنه من الصعوبة بمكان معرفة ذلك بدون إجراء العمليات الحسابية. إن هذا الأمر يمكن أن يكون محبطاً للقارئ الذي يشعر أن مستوى ألفا 0.01 تكون مسوغاً لهذه الأحوال أكثر من ألفا 0.05. وعند صياغة الاحتمالية الفعلية لدرجة عينة Z، بمعنى آخر، تعطي القراء المعلومات القصوى، بحيث يمكنهم الوصول إلى النتائج المتعلقة بهم حول ما إذا كانوا رافضين للفرض الصفري أو غير رافضين له، مفترضين استعدادهم للوقوع في الخطأ من النوع الأول أو الخطأ من النوع الثاني.

لقد تبين لنا أنه يمكننا الوصول إلى واحد من القرارين حول الفرض الصفري الذي مفاده (لافرق): نرفضه أو لا نرفضه. في أي من الحالتين، نحتاج إلى أن نساءل أنفسنا ما إذا كنا قد برهنا على أي شيء. فالإجابة بلا ! يجب أن نتذكر أننا نتعامل مع الاحتماليات فقط، ونحن ببساطة نقرر ما إذا كان الفرض الصفري جديراً بالقبول مقارنة بنتيجة العينة. فالعينات ليست دائماً هي المرآة الحقيقية للمجتمعات التي تُسْحَبُ منها هذه العينات، وبقيامنا بالاستدلال من العينة إلى المجتمع دائماً ينطوي على خطر الخطأ. آخذين هذه

النقطة العامة في الاعتبار حول القرارات التي يمكن أن نصل إليها، يتطلب منا أن نستكشف المعنى المحدد لهذين القرارين:

ماذا يعني قرار العجز في رفض الفرض الصفري؟

بالرغم من أننا نبدأ بالفرضية التي مفادها أن الفرض الصفري يكون فرضاً صحيحاً وبعد ذلك نستمر في اختبار هذه الفرضية، فالباحثون عادة يرغبون في رفض الفرض الصفري. عادةً وبشكل طبيعي نقوم بالمقارنة لأننا نعتقد بوجود فرق (نريد أن تكون القيمة الاحتمالية منخفضة)، إن قرار رفض الفرض الصفري هي النتيجة المرغوب فيها. بمعنى آخر، إننا نستخدم منطق البرهان الذي يقودنا إلى دعم الفرض البديل ومن خلال بيان أنه لا توجد معلومات داعمة لنقيضه، الفرض الصفري.

هل هذا يعني إذا عجزنا عن رفض الفرض الصفري، فالفرق الذي نبحث عنه غير موجود؟ للإجابة على ذلك، إنه ليس بالضرورة: أن العجز في رفض الفرض الصفري – لا يوجد فرق – يعني ببساطة لا توجد لدينا معلومات كافية للتفكير في أن الفرض الصفري فرض غير صحيح. وهذا لا يعني بالضرورة – على أية حال – أننا على صواب. ربما في الحقيقة يوجد فرق ولكنه على أساس أن مثل هذا الفرق لم يتبين من خلال نتيجة العينة. وهذا شيء يشبه افتراض البراءة في قانون الجريمة: فالمدَّعَى عليه أن يفترض براءته حتى تتوفر كل القرائن الكافية والقوية ليثبت حكم إدانته. وعلى أية حال، عندما نجد شخصاً ما بأنه ليس مذنباً استناداً على توفر قرائن قوية، هذا لا يعني، بالضرورة أن الشخص هو في حقيقة الأمر بريء: إن كل ما نعنيه هو القول بأن أيّاً من هذه الأحكام ممكنة، باعتبار أننا لا نختار حكم الإدانة إلا بعد توفر قرائن قوية للحقيقة. وبنفس الكيفية بالنسبة لحكم لا فرق، فإن عجزنا في عدم رفض الفرض الصفري لا يعني أن الفرض البديل على خطأ، إنما يعني ببساطة، استناداً على المعلومات المتوفرة فإن الفرض الصفري يستطيع أن يفسر نتائج العينة بدون أن نوسع فكرتنا الاحتمالية الممكنة والمعقولة.

إن العجز في إيجاد الفرق الدال يجب أن لا ننظر إليه كعملية مقنعة ونهائية. فإذا

توفرت لدينا خلفية نظرية جيدة للشك أو الارتياب بأن هذا الفرق لا وجود له بالرغم من أن الاختبار افترض بعدم وجوده. إن هذه القضية ربما تكون أساساً لبحوث مستقبلية. كذلك يمكننا القول بأن المتغير لم يخضع للتعريف الإجرائي بشكل صحيح. أو أن مستوى القياس لم يكن ليوفر لنا معلومات كافية. أو أن العينة لم يتم اختيارها بشكل ملائم أو لم تكن كبيرة.

وتجدر الإشارة إلى أنه في سياق البحوث الاجتماعية، بأن اختبارات الدلالة لا تبرهن على أي شيء باعتبارها شواهد للنقاش أو مناظرة قلما تصل فيها إلى نتائج حاسمة.

ماذا يعني قرار رفض الفرض الصفري؟

ماذا يكون الأمر لو كان قرارنا عكسيّاً: أي رفض الفرض الصفري بلغة الإحصاء يمكننا القول أن هناك فرقاً ذا دلالة إحصائية. ولكن ماذا يعني هذا؟ ما الذي تعلمناه حول العالم الواقعي، وهل يتوجب علينا عمل أي شيء حوله. إن مثل هذين السؤالين ليس بمقدور الاختبار الإحصائي الإجابة عليهما. إن الفرق الإحصائي ببساطة هو فرق بين عددين. طالب قد تحصل على درجة 60 في امتحان ما تختلف درجته إحصائياً عن طالب آخر تحصل على درجة 61 أو 66 أو 90. وبنفس التماثل يمكن أن يكون هناك فرق إحصائي ذا دلالة بين عينة من الطلاب يصل متوسط درجاتهم في امتحان نهائي 60 درجة وعينة أخرى. يصل متوسط درجاتها إلى 61 أو 66 أو 90. إن الفرق الإحصائي الدال ببساطة يخبرنا أن الرقمين ليسا متساويين. إن السؤال المطروح هنا هو: ما إذا كان مثل هذا الفرق يحتوي على أي أهمية نظرية أو عملية. ما إذا كان دالاً بما تحمله هذه الكلمة من معنى، للإجابة على هذا التساؤل إنه في حقيقة الأمر يمكننا كباحثين أو صانعي سياسات أن نقرر ذلك لأنفسنا.

ولتوضيح ذلك كله، دعنا نعطي مثالاً تطبيقياً حياً نفترض أن محاضراً لمادة الحاسب الآلي بالجامعة يود معرفة ما إذا كانت الجامعة على استعداد لزيادة الصرف المالي على ورش العمل المتعلقة بالحاسب الآلي. ومن خلال تعيين محاضرين إضافيين لمساعدة

الطلاب في استيعاب هذه المادة. فقد ترى الجامعة أنه يمكن أن تقوم بذلك الأمر فقط إذا ما تبين لها أن هناك فرقاً ذا دلالة بين درجات الطلاب في مواد الحاسوب ودرجاتهم في المواد الأخرى داخل الجامعة. قام المحاضر بجمع عينة من الطلاب لإيجاد متوسط درجاتهم في مادة الحاسوب التي وصلت إلى 55 درجة، وبمقارنة هذا المتوسط بمتوسط كل المواد داخل الجامعة الذي وصل إلى 62 درجة، ومن خلال هذا وجد أن الفرق دال إحصائياً على مستوى ألفا 0.05. السؤال الذي يمكن طرحه هنا هو: هل يمكننا أن نعتقد أن المحاضر بهذا العمل قد أقنع الجامعة؟ ليس بالضرورة أن المحاضر يمكن أن يضع في اعتباره أن الفرق في متوسط الدرجات يبرر المصروفات المالية الإضافية بسبب اعتقاده أن مادة الحاسوب مادة مهمة للعملية التعليمية برمتها في الجامعة. إلا أن الجامعة تمتلك الحق في القول بأن كل الطرق الافتراضية الممكنة تجعلها تقوم بعملية الصرف المالي، لكن الفرق المتمثل في سبع درجات شيء يمكن أن يقبل. فالجامعة - على الجانب الآخر - قد لا يكون لديها أي حجة مع المحاضر حول الفرق الإحصائي الدال؛ أي أن إدارة الجامعة قد قبلت أن هناك فرقاً في الدرجات داخل الجامعة لا يرجع في الحقيقة إلى تباين المعاينة. إلا أن هذا - على أية حال - قد يحث إدارة الجامعة للصرف كي تحاول ردم الهوة.

إن هذا المثال يوضح لنا في الغالب أن هناك نقطة تم تجاهلها. وهي أنه ليس من الشائع للباحثين، بكل بساطة، أن يبينوا أن نتيجة ما دالة على مستوى 0.05 أو 0.01 بدون ملاحظات إضافية، وكأن هذا كل ما نحتاج قوله. في حقيقة الأمر، يجب أن تكون هذه بداية لإبداعات واهتمامات بحثية أكثر. ما الذي يمكننا أن نقرره حول العالم الواقعي؟ وما الذي يمكننا فعله من خلال رفض الفرض الصفري؟ [5].

اختبار ثنائي الجانب لـ Z لمتوسط مفرد:

نفترض أن الجامعة ترغب في معرفة متوسط الأداء الأكاديمي للطلاب الوافدين في استيعابهم للمسارات الدراسية داخل الجامعة. إن الجامعة على علم بأن متوسط درجات الطلاب يصل إلى 62 درجة، بانحراف معياري 15 درجة. ومن عينة عشوائية لعدد 150 طالباً وافداً تم سحبها عشوائياً وصل متوسطها الحسابي إلى 60.5 درجة.

الإجراءات:

الخطوة الأولى: صياغة الفروض:

هل متوسط درجات الطلاب الوافدين يختلف عن متوسط باقي طلاب الجامعة؟ من خلال هذا السؤال البحثي، يمكننا صياغة فرضين أساسيين:

- الفرض الصفري Ho: لا يوجد فرق بين متوسط درجات الطلاب الوافدين، ومتوسط باقي طلاب الجامعة.

$$Ho: u = 62$$

- الفرض البديل: يوجد فرق بين متوسط درجات الطلاب الوافدين وباقي متوسط درجات طلاب الجامعة.

$$Hi: u \neq 62$$

الخطوة الثانية: اختيار درجة الدلالة:

إن العامل المهم في هذا المثال، أننا نرغب في معرفة المتوسط، وأن المتغير هو الأداء الأكاديمي الذي تم قياسه على مستوى القياس ذي المسافات والنسبي (الدرجة النهائية). إن دور الإحصاء الوصفي هنا، هو مساعدتنا في حساب تلخيص البيانات المتعلقة بالمتوسط الحسابي. إن هذين العاملين يسمحان لنا بإجراء اختبار Z لمتوسط مفرد (One Single Mean).

الخطوة الثالثة: حساب درجات العينة:

من المعلومات الواردة في هذا المثال، يمكننا إجراء الاختبار الإحصائي:

$$Z = \frac{\overline{x} - u}{\frac{\sigma}{\sqrt{N}}} = \frac{60.5 - 62}{\frac{15}{\sqrt{150}}} =$$

$$\frac{-1.5}{1.22} = -1.2$$

$$Z = -1.2$$

الخطوة الرابعة: بيان المنطقة الحرجة أو منطقة الرفض (ألفا $\alpha = 0.05$)

في صياغتنا للفرض البديل: قلنا إن هناك فرقاً بين متوسط الطلاب الوافدين وباقي متوسط طلاب الجامعة $Hi: u \neq 62$.

لاحظ أننا في صياغة هذا الفرض لم نتطرق البتة إلى القول بأن متوسط درجات الطلاب الوافدين ستكون أعلى أو أقل من متوسط درجات الطلاب الآخرين. ومن هنا جاء اختبار ثنائي الجانب، لأنه لا يوجد لدينا أي سبب للقول بأن الطلاب الوافدين أحسن أو أسوأ من الطلاب الآخرين وفقاً لمتوسط الدرجات.

من خلال توزيعات المساحة تحت المنحنى الطبيعي المعياري، فالدرجات الحرجة لاختبار ثنائي الجانب بالفا $\alpha = 0.05$ هي

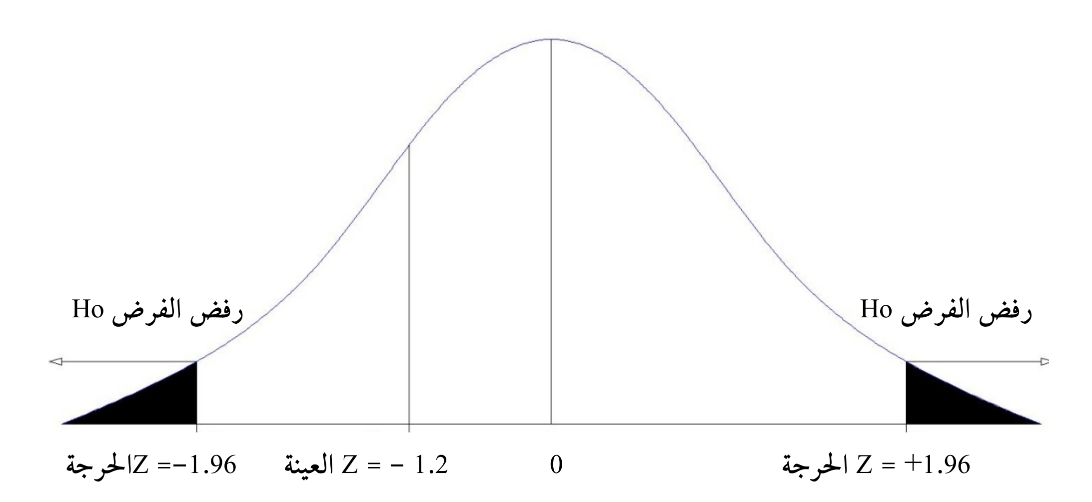
$$Z_{\text{الحرجة}} = \pm 1.96$$

الخطوة الخامسة: اتخاذ القرار من خلال المعلومات التي تحصلنا عليها:

$$Z_{\text{العينة}} = -1.2$$

$$Z_{\text{الحرجة}} = \pm 1.96$$

وعليه لا يمكننا رفض الفرض الصفري الذي مفاده: لا يوجد فرق بين متوسط درجات الطلاب الوافدين ومتوسط درجات باقي طلاب الجامعة فيما يتعلق بموضوع الأداء الأكاديمي.

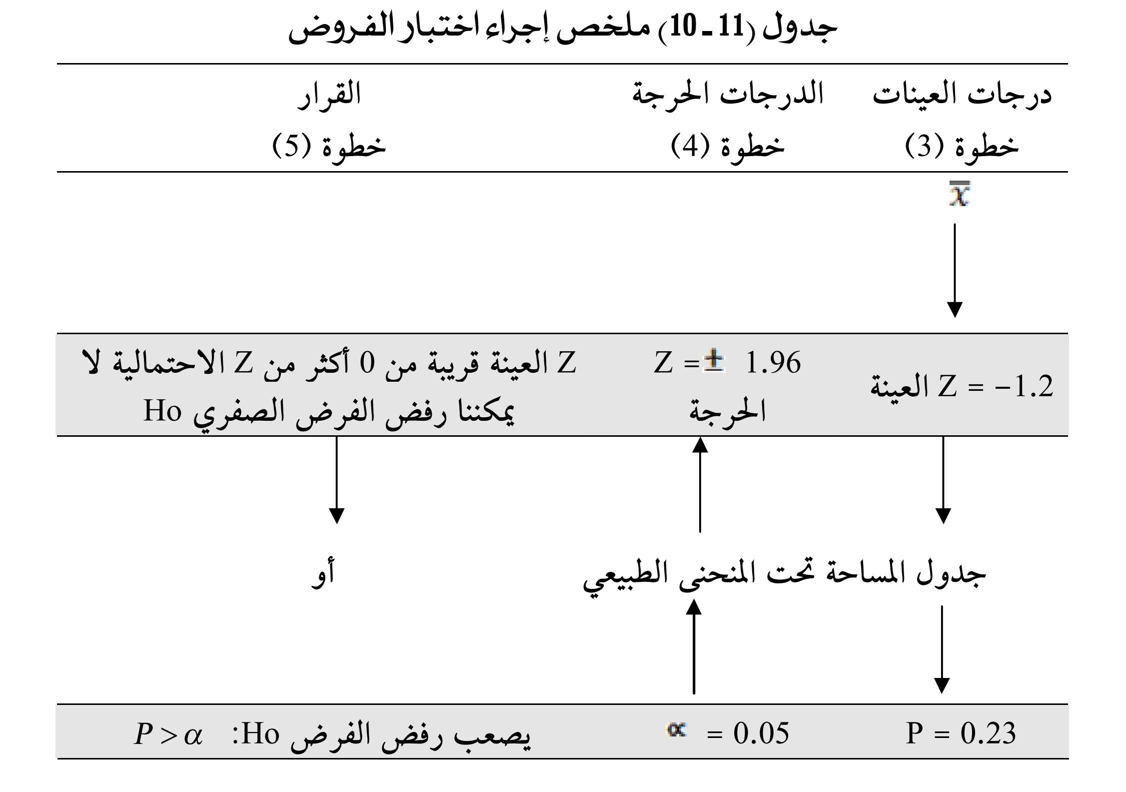


شكل (11 ـ 4) درجات العينة والدرجات الحرجة

يمكننا القول هنا بأن الفرق الإحصائي يصل إلى 1.5 بين درجات عينة الطلاب الوافدين، والقيمة المفترضة هي ببساطة راجعة إلى التباين العشوائي عند سحب العينة من المجتمع المدروس الذي وصل متوسط درجاته إلى 62 درجة.

كذلك يمكننا الوصول إلى قرار بمقارنة الاحتماليات المرتبطة بهذه العينات، والقيم الحرجة. إن درجة الدلالة التي تم اختيارها هي $\alpha = 0.05$. وبالرجوع إلى جدول التوزيع الطبيعي لدرجات Z التي تساوي 1.2−، من خلال ذلك يمكننا القول عند معاينة جدول المساحة تحت المنحنى أبعد من كلتا النقطتين، فإن قيمة (The P value) هي 0.23. بمعنى آخر، إذا حصل مجتمع الطلاب الوافدين على متوسط درجات 62، فإن 23 عينة في كل مائة سحبت من هذا المجتمع الذي سيصل متوسط درجاته إلى 1.5 أو أبعد من 62. انظر جدول رقم (11 − 10).

جدول (11ـ10) ملخص إجراء اختبار الفروض



اختبار Z أحادي الجانب لمتوسط مفرد:

مثال: مجموعة من المنتجين في مصنع ما قد شكوا بأن الأوضاع المهنية غير آمنة، مما سبب لهم معدلاً عالياً من ضيق التنفس Respiratory Illness مما جعلهم يطلبون من كلية الصحة العامة أن نقوم بإجراء دراسة حولهم. من خلال اختيار عينة عشوائية من 100 منتجٍ. وطرح السؤال التالي على كل منتج من هؤلاء المنتجين مفاده: كم عدد الساعات التي فقدها المنتج في السنة كنتيجة لمرض ضيق التنفس؟ إن متوسط عدد الساعات المفقودة وصل إلى 15 ساعةً في الأسبوع. ومن المعلومات المتوفرة من المصادر الرسمية فإن الباحث يعي أن كل المنتجين بالمصانع في المجتمع يصل متوسط عدد الساعات المفقودة منهم نتيجةً لمرض ضيق التنفس إلى 12 ساعة في العام، بانحراف معياري يصل إلى 7.5 ساعة حول المتوسط.

لقد جادل المنتجون في أن ما أوضحته نتائج العينة المسحوبة هو أن معدل مرض ضيق التنفس أعلى منه لدى المنتجين الآخرين، إلا أن إدارة المصنع قد أدعت أن الفرق بين معدلات مرض التنفس لدى العينة المدروسة ومعدلات المنتجين الآخرين هو فرق صغير جداً. يرجع في الأساس إلى التباين العشوائي عند المعاينة. وبكل وضوح فقد توجد بعض الفروق بين عينة المنتجين في المصنع وباقي أفراد المصانع الأخرى، ولكن هذا الفرق بدرجة كبيرة وكافية يقترح بأن هناك أكثر من مجرد احتمالية الصدفة. إن هذه الإدعاءات يمكن اختبارها:

الإجراءات:

الخطوة الأولى: صياغة الفروض:

- الفرض الصفري: لا يوجد فرق بين معدل مرض ضيق التنفس الذي يعاني منه المنتجون في هذا المصنع والمعدل الذي يعاني منه المنتجون في كل المصانع.

Ho: u = 12 ساعة

- الفرض البديل: يوجد فرق بين معدل مرض ضيق التنفس الذي يعاني منه المنتجون في هذا المصنع والمعدل الذي يعاني منه المنتجون في المصانع ما عدا أن هؤلاء المنتجين في هذا المصنع لديهم معدل عالٍ من هذا المرض.

Hi: u > 12 ساعة

لاحظ أن الفرض البديل لا يبين الفرق فقط وإنما اتجاه هذا الفرق. فالمنتجون لديهم الرغبة فقط في رفض الفرض الصفري Ho إذا تبين أنهم يسجلون معدلات عالية لهذا المرض ليكون مبرراً كافياً للمطالبة بالتعويض، وبالتالي أصبح من الضرورة بمكان تحديد ما إذا كنا نريد أن نجري اختبار أحادي الجانب أم ثنائي الجانب "الخطوة الرابعة".

الخطوة الثانية: اختيار اختبار درجة الدلالة:

في هذا المثال نحن نرغب في معرفة متوسط مرض ضيق التنفس لدى هؤلاء المنتجين

والذي تم قياسه على مستوى ذي المسافات والنسبي "عدد الساعات التي فقدت" قد تم استخدام الإحصاء الوصفي الملائم لتلخيص البيانات التي تحصل عليها الباحث من خلال طرحه لسؤال بحثي وهو المتوسط: وهذه البيانات ستكون صالحة لإجراء اختبار أحادي الجانب لمتوسط مفرد Z.

الخطوة الثالثة: إجراء العمليات الحسابية:

إن المعلومات ذات الصلة التي من خلالها نستطيع حساب درجة العينة هي

$$N = 100$$

$\sigma =$ 7.5 يوم

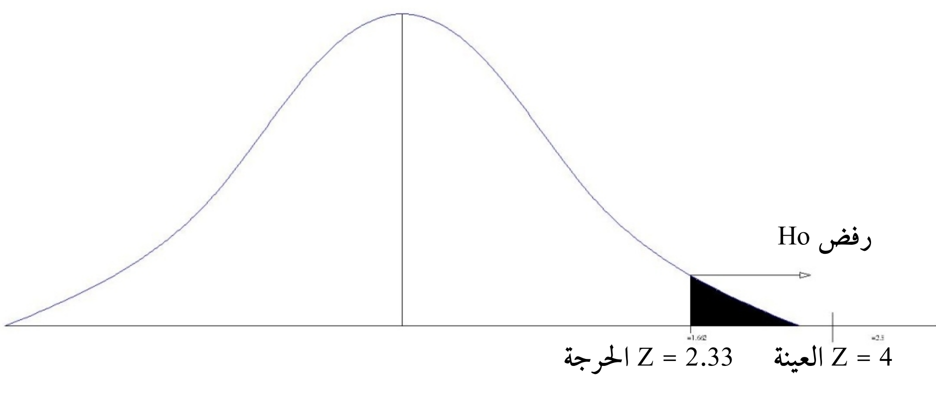
$\overline{X} =$ 15 يوم

$$Z = \frac{\overline{x} - u}{\frac{\sigma}{\sqrt{N}}} = \frac{15 - 12}{\frac{7.5}{\sqrt{100}}} = 4$$

وبالنظر إلى الجدول المتعلق بالمساحة تحت المنحنى الطبيعي نجد أن درجة الدلالة المرتبطة بدرجة Z للعينة المدروسة هي $P < 0.0001$.

الخطوة الرابعة: إيجاد الدرجات الحرجة والمنطقة الحرجة:

على الإدارة أن تعطي من المبررات ما يقنع المنتجين بأن هذا الفرق العالي في معدل مرض ضيق التنفس لهذا المصنع، أنه من المرجح أن يكون هذا الفرق هو نتيجة لاختيار العينة. ومن هنا تم اختيار 0.01 كمستوى لدلالة، ولما كان الفرض البديل لا يحدد الفرق فقط، بل أيضاً اتجاه هذا الفرق الأمر الذي يدعونا إلى استخدام المنحنى الطبيعي نتحصل على الدرجات الحرجة لـِ Z وهي 2.33. إن اتجاه الفرق الذي تبين في الفرض البديل سيكون متوسطه أكبر من 12 ساعة. وعليه فإن الجانب المناسب هو الجانب الأيمن من التوزيع:



شكل (11ـ5) المنطقة الحرجة ودرجات العينة

الخطوة الخامسة: اتخاذ القرار:

إذا ما تمت مقارنة الدرجة المحسوبة لـِ Z مع الدرجة الحرجة لـِ Z انظر الشكل أعلاه، فإننا في هذه الحالة نرفض الفرض الصفري باعتبار أن الدرجة المحسوبة لـِ Z أكبر من الدرجة الاحتمالية ومن هنا يصبح لدى المنتجين الحجة المشروعة للمطالبة بالتعويض على هذا المرض(6).

في هذا الفصل بينا الخطوات الأساسية الأكثر ارتباطاً بإجراءات الفروض إلا أن هناك بعض المحاذير المتعلقة باختبار الفرض يمكننا التطرق لها في هذا السياق.

المحاذير المتعلقة باختبار الفرض: قياس حجم التأثير:

بالرغم من شيوع استخدام تقنيات اختبار الفرض لتقييم وتفسير البيانات، إلا أن هناك بعضاً من العلماء قد أثاروا جملة من التحفظات حول إجراءات اختبار الفروض. ولعل أكثر هذه الانتقادات حدة تكمن في تفسير النتيجة الدالة A Significant result.

في حقيقة الأمر، هناك نوعان من القيود الجدية في استخدام اختبار الفرض لتأسيس تأثير دلالة المعالجة، إن أول هذه القيود في اختبار الفرض تركز على البيانات أكثر من

الفرض ذاته. لاسيما عندما يكون الفرض قد تم رفضه، فإننا في حقيقة الأمر، نصل إلى بيان احتمالي قوي حول بيانات العينة، وليس حول الفرض الصفري.

إن النتيجة الدالة تسمح بالنتيجة التالية: أن متوسط العينة المحدد يكون جداً بعيد الاحتمال ($P < .05$) إذا كان الفرض الصفري فرضاً صحيحاً. لاحظ أن هذه النتيجة لا تعطينا بياناً واضحاً حول الاحتمالية بأن يكون الفرض الصفري فرضاً صحيحاً أو خطأ. فالحقيقة أن البيانات تكون بعيدة الاحتمال لنقترح أن الفرض الصفري يكون أيضاً بعيد الاحتمال، ولكنه لم يكن لدينا أي خلفية قوية لنصل إلى بيان احتمالي حول الفرض الصفري. وبرفض الفرض الصفري بمستوى دلالة $\alpha = 0.05$ لا يبرر النتيجة الاحتمالية بأن الفرض الصفري يكون فرضاً صحيحاً بأقل من 5 %. وثاني هذه القيود يتعلق بالبرهنة بأن تأثير المعالجة الدالة لا يشير بالضرورة إلى تأثير معالجة حقيقي. وجدير بالذكر، أن الدلالة الإحصائية لا تمدنا بأيّة معلومات حقيقية حول الحجم المطلق لتأثير المعالجة. وبدلاً من ذلك، فإن اختبار الفرض يؤسس ببساطة، بأن النتائج المُتَوَصَّل إليها في الدراسة البحثية تكون بعيدة الاحتمال عن الحدوث إذا لم يكن هناك تأثير معالجة. ويصل اختبار الفرض لهذه النتيجة من خلال:

1- حساب الخطأ المعياري The Standard Error الذي يقيس كم هو حجم الفرق الذي يكون معقولاً وعلى نحو متوقع بين $\overline{X}$ و u.

2- البرهنة على أن فرق المتوسط المتحصل عليه هو فعلياً أكبر من الخطأ المعياري. تجدر الإشارة هنا إلى أنه إذا كان الخطأ المعياري صغيراً جداً، حينئذٍ، فإن تأثير المعالجة يمكن أن يكون أيضاً هو الآخر صغيراً جداً مع البقاء بأنه كبير بشكل كافٍ لأن يكون دالاً؛ وبالتالي فإن تأثير الدلالة لا يعني بالضرورة أنه تأثير كبير.

ولتوضيح فكرة أن اختبار الفرض يقيم الحجم النسبي لتأثير المعالجة أكثر من تقييم الحجم المطلق، يمكن بيانه في المثال التالي:

مثال: دعنا، نبدأ بدرجات مجتمع يشكل توزيعاً طبيعياً لـِ: $u = 50$ و $\sigma = 10$. وقد تم اختيار عينة من هذا المجتمع وتمت معالجة هذه العينة. وبعد المعالجة تحصلنا على

متوسط لهذه العينة يساوي: $\overline{X} = 51$. السؤال المطروح هنا هو: هل هذه العينة تقدم الدليل بأن تأثير المعالجة يكون دالاً إحصائياً؟ بالرغم من أن هناك فرق نقطة واحدة بين متوسط العينة ومتوسط المجتمع الأصلي، إلا أن هذا الفرق يمكن أن يكون كافياً لأن يكون فرقاً دالاً. وجدير بالذكر، أن نتائج الفرض نتائج تعتمد على حجم العينة. فعلى سبيل المثال، إذا كانت لدينا عينة $N = 25$، والخطأ المعياري:

$$\sigma \overline{X} = \frac{\sigma}{\sqrt{N}} = \frac{10}{\sqrt{25}} = \frac{10}{5} = 2.00$$

وأن درجة Z لِـ $\overline{X}$ تساوي $\overline{X} = 51$ فإن:

$$Z = \frac{\overline{x} - u}{\sigma \overline{x}} = \frac{51-50}{2.00} = \frac{1}{2} = 0.50$$

وبمستوى دلالة $\alpha = .05$ فإن المنطقة الحرجة ستبدأ عند $Z = 1.96$. ودرجة Z فشلت في أن تصل إلى المنطقة الحرجة؛ لذلك لا يمكننا رفض الفرض الصفري. وفي هذه الحالة، فإن "فرق النقطة الواحدة" بين $\overline{x}$ و u ليست دالة بسبب أنها قدرت نسبة إلى الخطأ المعياري بدرجتين.

دعنا نتعامل مع عينة $N = 400$، والخطأ المعياري:

$$\sigma \overline{X} = \frac{\sigma}{\sqrt{N}} = \frac{10}{\sqrt{400}} = \frac{10}{20} = 0.50$$

وأن درجة Z لِـ $\overline{x}$ تساوي $\overline{x} = 51$ فإن:

$$Z = \frac{\overline{x} - u}{\sigma \overline{x}} = \frac{51-50}{0.50} = \frac{1}{50} = 2.00$$

الآن درجة Z أبعد من حدود 1.96. وعليه، نرفض الفرض الصفري، ونصل إلى نتيجة مفادها أن هناك تأثيراً دالاً. في هذه الحالة، فإن "فرق النقطة الواحدة" بين $\overline{x}$ و u اعتبرت دالة بسبب أنها قُدّرت نسبة إلى الخطأ المعياري فقط بـِ: 0.50 نقطة.

تجدر الإشارة من خلال هذا المثال إلى أنه يمكن أن يبقى تأثير المعالجة الصغير دالاً إحصائياً. فإذا كان حجم العينة كبيراً بشكل كافٍ، فإن تأثير أي معالجة مهما كان صغره، يمكن أن يكون كافياً لنا لنرفض الفرض الصفري.

قياس حجم التأثير Measuring Effect Size:

كما بينا أعلاه، بالنسبة لأحد هذه القيود المتعلقة باختبار الفرض، أن اختبار الفرض لا يقيم حقاً الحجم المطلق لتأثير المعالجة. ولتصحيح هذه المعضلة، ينبغي على الباحث عند وصفه بأن هناك تأثيراً دالاً إحصائياً أن يبين أيضاً وصفاً يتعلق بحجم التأثير Effect Size. ويعني حجم التأثير توفير القياس المطلق لتأثير المعالجة، مستقلاً عن حجم العينة التي يتم استخدامها.

إن أسهل، وأكثر طريقة لقياس حجم التأثير هي طريقة كوهينز d: Cohenś d من خلال المعادلة التالية:

$$\text{Cohenś d} = \frac{\text{فرق المتوسط}}{\text{الانحراف المعياري}}$$

وكبرهنة نهائية واحدة لـِ Cohenś d، يمكننا الرجوع إلى المثال السابق لكل واحد من الاختبارين. لقد كان المتوسط الأصلي للمجتمع $u = 50$، بانحراف معياري $\sigma = 10$، لكل اختبار، والمتوسط للعينة المعالجة كان $\bar{x} = 51$. وبالرغم من أن أحد الاختبارين قد استخدم عينة $N = 25$، والاختبار الآخر استخدم عينة $N = 400$، فإن حجم العينة لم يُؤخذ في الاعتبار عند حساب Cohenś d؛ عليه، فإن كلا الاختبارين يولدان نفس القيمة:

$$\text{Cohenś d} = \frac{\text{فرق المتوسط}}{\text{الانحراف المعياري}} = \frac{1}{10} = 0.10$$

وبمراجعة الجدول التالي، يمكننا القول أن Cohenś d يصف ببساطة حجم تأثير المعالجة، ولم يتأثر البتة بعدد الدرجات في العينة. ولكلا الاختبارين، فإن المتوسط الأصلي كان 50، وبعد المعالجة كان المتوسط 51؛ وبالتالي تظهر المعالجة أنها زادت الدرجات بدرجة واحدة مساوية لواحد من عشرة من الانحراف المعياري One-tenth (كوهينز d تساوي 0.1) [7].

حجم d	تقييم حجم التأثير
d = 0.2	تأثير صغير (فرق المتوسط يدور حول 0.2 انحراف معياري)
d = 0.5	تأثير متوسط (فرق المتوسط يدور حول 0.5 انحراف معياري)
d = 0.8	تأثير عال (فرق المتوسط يدور حول 0.8 انحراف معياري)

أسئلة للمراجعة:

1- تحت أي الظروف سيكون متوسط العينات لتوزيع المعاينة يتخذ شكلاً طبيعياً؟

2- ماذا نعني بالخطأ من النوع الأول Type I Error، والخطأ من النوع الثاني Type II Error، وما هي العلاقة بينهما؟

3- بين كيف أن اختيار مستوى الدلالة يؤثر على المنطقة الحرجة؟

4- أكمل الجدول التالي:

الاحتمالية	الاختبار	درجة Z
0.230	....	± 1.2
0.100	ثنائي الجانب	....
0.018	....	± 2.1
....	ثنائي الجانب	± 2.3
....	أحادي الجانب	± 3.4

5- ضع مخططاً للمنطقة الحرجة للدرجات الحالية التالية:

أ- $Z > 1.645$ ب- $Z < -1.645$ ج- $Z > 1.96$

أو $Z < -1.96$

ما هي احتمالية الخطأ من النوع الأول التي ترتبط بهذه المناطق الحرجة؟

6- من البيانات التالية أحسب درجة Z:

N	$\bar{x}$	σ	U	
180	2.3	0.7	2.4	أ-
100	16.7	1.1	18	ب-

7- عينة بمتوسط حسابي 12 سنة تم اختبارها ما إذا كانت قد سحبت من مجتمع بمتوسط حسابي يصل إلى 15 سنة:

أ- درجة الدلالة لاختبار ثنائي الجانب برهنت على أن تكون 0.03. اشرح بلغة بسيطة إلى ما يشير هذا؟

ب- درجة الدلالة لاختبار أحادي الجانب هي 0.015. اشرح بلغة بسيطة إلى ما يشير هذا؟

8- اشرح باختصار لماذا يعتبر أن تأثير الدلالة الإحصائية ليس بالضرورة تأثيراً حقيقيّاً؟

9- احسب قيمة Cohenś d من بيانات بينت أن هناك فرق المتوسط يساوي 15 نقطة. قبل المعالجة كان متوسط الدرجات 500، وبعد المعالجة كان متوسط الدرجات 515. والانحراف المعياري 100 = σ .

10- إذا كانت u = 65 وقيمة 70 = $\bar{x}$ و 15 = σ.

بين فرق المتوسط وقيمة Cohenś d. وماذا تعني لك هذه القيمة المحسوبة.

الهوامش والمصادر:

أولا: الهوامش:

1- George Argyrous , Statistics for Social and Health Research , With a Guide to Spss Sage Publications ,USA , 2001 , PP. 262 - 267.

2- انظر: عبد الله عامر الهمالي، التقنيات الإحصائية ومناهج البحث: مدخل نظري وتطبيقي للعلوم الاجتماعية، منشورات جامعة قاريونس - بنغازي، 2008، ص 106.

3- المصدر نفسه، ص ص 108 - 112.

4- George Argyrous , op.cit., PP. 272 - 273.

5- Ibid , PP. 274 - 280.

6- Ibid , PP. 280 - 285.

7- Frederick J. Gravetter and Larry B. Wallnau, Statistics for The Behavioral Sciencies, 8th ed, Wadsworth Cengage Learning, USA, 2010, PP. 260-264.

ثانيا: المصادر:

1- Frederick J. Gravetter and Larry B. Wallnau , Statistics for The Behavioral Sciencies , 8th ed , Wadsworth Cengage Learning , USA , 2010.

2- George Argyrous , Statistics for Social and Health Research , With a Guide to Spss , Sage Publications ,USA , 2001.

3- Joseph F. Healey , The Essentials of Statistics: A TooL for Social Research , Wadsworth Cengage Learning , UAS , 2010.

4- عبد الله عامر الهمالي، التقنيات الإحصائية ومناهج البحث: مدخل نظري وتطبيقي للعلوم الاجتماعية، منشورات جامعة قاريونس - بنغازي، 2008.

الفصل الثاني عشر

اختبار t لمتوسط حسابي لعينة واحدة

مقدمة:

تناولنا في الفصل السابق منطق اختبار الفرض الإحصائي، وإن المتتبع لإجراءات اختبار Z لمتوسط عينة واحدة يلاحظ أننا استخدمنا الانحراف المعياري للمجتمع من أجل إجراء الاستدلال من متوسط العينة على متوسط المجتمع، أيضاً يمكن للقارئ الناقد أن يفكر في هذا الموقف: أن البيانات التي تم استخدامها لحساب الانحراف المعياري للمجتمع يجب أيضاً أن تسمح لنا بشكل مباشر بحساب المتوسط. وإذا كان الانحراف المعياري للمجتمع لدينا معروفاً فالسؤال: كيف لا نستطيع أن نعرف متوسط المجتمع؟ تجدر الإشارة إلى أنه لا توجد حاجة إلى أن نقوم بالاستدلال من العينة على متوسط المجتمع، ولكن بإمكاننا أن نكون بشكل مباشر قادرين على حسابه، بمعنى آخر، أنه من غير المحتمل أن نواجه موقفاً لا نعرف فيه الانحراف المعياري للمجتمع لكننا قد لا نعرف متوسطه. وإذا كنا قد بدأنا في الفصل السابق باختبار Z لمتوسط عينة واحدة لأنه اختبار يبين لنا وبشكل بسيط الإجراء المتعلق باختبار الفرض وبعد أن تبين لنا وبشكل جلي الخطوات الأساسية للاختبار، الآن يمكننا الانتقال إلى تطبيق هذه الخطوات في مواقف أكثر تعقيداً.

إن الاختبارات التي سنتناولها في الفصول اللاحقة لهذا الكتاب هي اختبارات في مجملها متباينة في الإجراء الأساسي لاختبار الفرض وسوف نتعلم خلال هذا الفصل الشروط المناسبة لكل اختبار من هذه الاختبارات، وهي تلك العوامل التي سننظر إليها في الخطوة الثانية من إجراء اختبار الفرض لنحدد دلالة الاختبار من أجل توظيفه.

ثمة عاملان أساسيان ينبغي وضعهما في الاعتبار: أولهما الإحصاء الوصفي الذي يستخدم في تلخيص بيانات العينة. وثانيهما عدد العينات التي حولها تم صياغة الفرض[1].

في هذا الفصل سيتم التركيز على اختبار t لمتوسط عينة واحدة والتي تستخدم بدلاً من اختبار Z لعينة واحدة في أكثر المواقف شيوعاً عندما يكون متوسط المجتمع والانحراف المعياري غير معروفين لدينا.

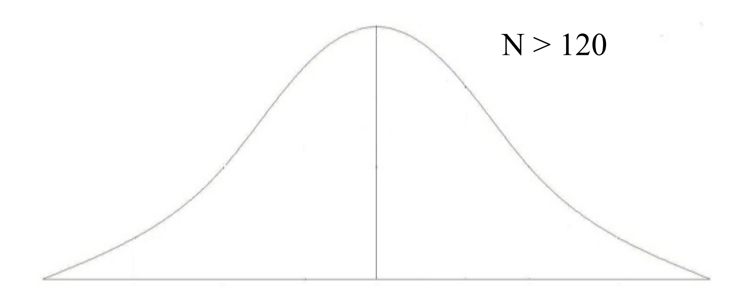
توزيع t:

عندما نريد أن نقوم باستدلال حول متوسط مجتمع ما حيث لا يكون لدينا معرفة بانحرافه المعياري فإن تغيراً طفيفاً مطلوب تعديله أو تغييره في الإجراءات الأساسية التي تمت مناقشتها في الفصل السابق. وعليه، فإننا لم نعد نستخدم توزيعات Z لاستنتاج درجات العينة والدرجات الحرجة ويرجع السبب في ذلك إلى أن توزيعات المعاينة لمتوسطات العينة لم تعد موزعة توزيعاً طبيعياً. وعوضاً عن ذلك يستخدم توزيع t.

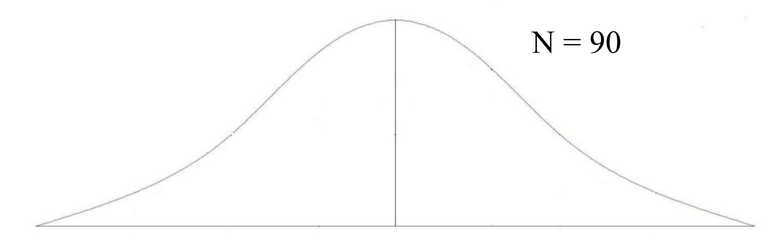
وبعد ذلك نجري اختبار t الذي يطلق عليه توزيع Student الذي يرجع الفضل فيه إلى W.Gossett.

إن توزيع t يشبه إلى حد كبير توزيع Z فهو توزيع سلس وأحادي ومنحنى متماثل، غير أن الفرق بينهما هو أن توزيع t أكثر تسطحاً مقارنة بتوزيع Z. والسؤال المطروح هو: كم بالضبط تكون هذه التوزيعات مسطحة؟ إن الإجابة على هذا السؤال يعتمد بالدرجة الأولى على حجم العينة (انظر الشكل التالي):

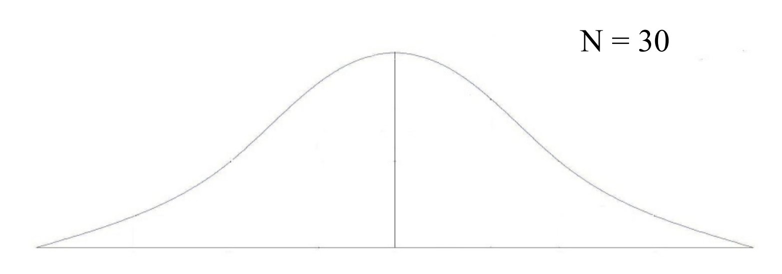
أ- توزيعات t عندما تكون:



ب- توزيعات t عندما تكون:



ج- توزيعات t عندما تكون:



شكل (12-1) توزيعات t لعينات تكون أحجامها

(أ) $N > 120$ ،(ب) $N = 90$ ،(ج) $N = 30$

إن توزيع t عندما يصل حجم العينة فيه إلى 30 فإن توزيعات t تكون ذات ديلين مسطحين وهذان الذيلان يصبحان رفيعين جداً عندما يكون حجم العينة 90 وهما في آخر الأمر متماثلان مع المنحنى الطبيعي عندما يكون حجم العينة أكبر من 120[2].

اختبار t لمتوسط حسابي لعينة واحدة:

في هذا الاختبار سنتعامل مع الخطوات الخمس المتعلقة بإجراء اختبار الفرض الإحصائي، مع ملاحظة الفرق بين هذه الإجراءات وتلك التي تم تطبيقها عند إجراء اختبار Z لمتوسط حسابي لعينة واحدة.

مثال: نفترض أن الهيئة الوطنية للمعلومات ترغب في معرفة ما إذا كان المتوسط العمري لسكان منطقة معينة هو أكثر من 40 سنة من أجل أن تقرر حجم الميزانية التي ستخصص لمستشفى في تلك المنطقة. ونظراً لصعوبة إجراء مسح شامل لهذه المنطقة عمدت الهيئة إلى أخذ عينة عشوائية من 51 مفردة من هذا المجتمع وصل متوسطها الحسابي إلى 43 سنة، بانحراف معياري 10 سنوات. ومن هنا نجد وبشكل واضح أن متوسط العينة أكبر من 40 سنة (متوسط المجتمع). وبناءً على هذه النتيجة وجدت الهيئة نفسها مترددة في القول بأن متوسط العينة التي سحبت من المجتمع يفوق الـ 40 سنة من العمر مما حدا بالهيئة إلى القول بأن العينة ببساطة جاءت من مجتمع بمتوسط عمري يصل إلى 40 سنة وأن تأثير التباين العشوائي هو المسئول عن تفسير الزيادة الطفيفة في المتوسط العمري للعينة. وبإمكاننا اختبار هذا الإدعاء مستخدمين اختبار t لمتوسط حسابي لعينة واحدة وفقاً للخطوات التالية:

الخطوة الأولى: صياغة الفرض الصفري والفرض البديل:

Ho: يصل المتوسط العمري لسكان هذا الإقليم إلى 40 سنة من العمر.

$$Ho: u = 40 \text{ سنة}$$

Hi: إن المتوسط العمري لسكان هذا الإقليم أكبر من 40 سنة.

$$Hi: u > 40 \text{ سنة}$$

لاحظ التفاوت في صياغة الفرض البديل. وبناءً على السياسة المعتمدة للهيئة فيما يتعلق بالتمويل فإننا لا نرغب في معرفة ما إذا كان سكان هذا الإقليم في المتوسط أصغر من 40 سنة لأن عملية التمويل سوف تتغير فقط، إذا ما وجدت الهيئة أن المتوسط العمري للسكان ذا دلالة أي أكبر من 40 سنة.

الخطوة الثانية: اختيار اختبار الدلالة:

ينصب التركيز هنا على المتوسط العمري لعينة واحدة وحيث إن العمر قد تم قياسه على مستوى ذي المسافات والنسبي فإن الإحصاء الوصفي المناسب هو المتوسط الحسابي، وعلى خلاف الأمثلة في الفصل السابق أنه ليست لدينا أية معلومات فيما يتعلق بالانحراف المعياري للمجتمع، وعليه فإننا سنستخدم اختبار t لمتوسط حسابي لعينة واحدة.

الخطوة الثالثة: حساب درجة العينة:

عندما يكون الانحراف المعياري للمجتمع غير معروف حينئذٍ يمكننا استخدام معادلة اختبار t بدلاً من معادلة اختبار Z لحساب درجة العينة.

هذه المعادلة ستعوض الانحراف المعياري للمجتمع بالانحراف المعياري للعينة:

$$t = \frac{\overline{x} - u}{\frac{s}{\sqrt{N-1}}} = \frac{43 - 40}{\frac{10}{\sqrt{51-1}}}$$

$$= 2.1$$

الخطوة الرابعة: تعيين الدرجة الحرجة والمنطقة الحرجة:

لتعيين الدرجة الحرجة والمنطقة الحرجة يتطلب الآتي:

أ- الاختيار بين اختبار أحادي الجانب واختبار ثنائي الجانب.

ب- اختيار مستوى الدلالة (مستوى ألفا) من خلال الفرض البديل الذي يحدد اتجاه الفرق.

هنا يمكننا استخدام اختبار أحادي الجانب (الديل الأيمن) واختبار ألفا بمستوى دلالة

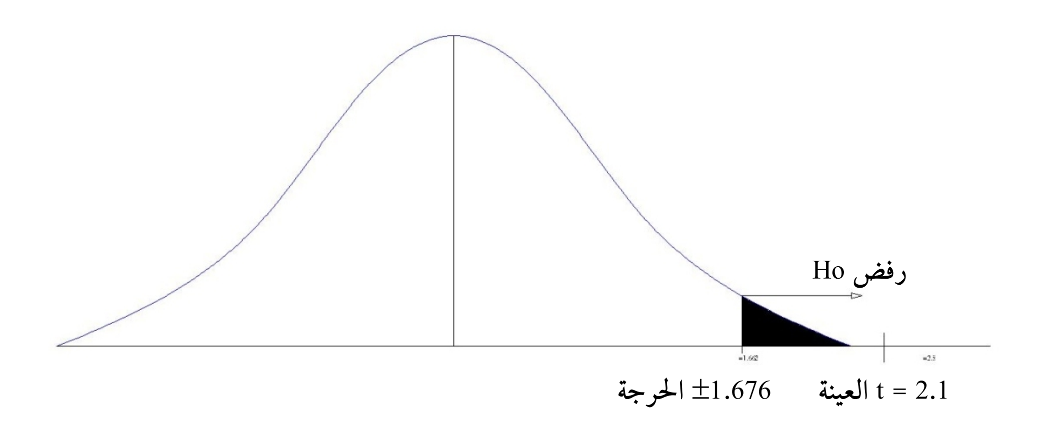
0.05 كما درج عليه في مجال البحث الاجتماعي. وعند اختيار مستوى ألفا وعدد الديلين، عندئذ يمكننا الرجوع إلى جدول القيم الحرجة لتوزيع t. وبما أن هناك فرقاً في توزيع t لكل حجم عينة في الجدول، فالجدول يزودنا بمجموعة القيم الحرجة ودرجات t لعدد مختلف من درجات الحرية df. ففي المثال الذي بين أيدينا فإن درجة الحرية تكون 50 أي:

$$Df = 51 - 1 = 50$$

وأن المنطقة الحرجة لاختبار أحادي الجانب بمستوى دلالة 0.05= α أي 1.676±، انظر شكل (12 – 2).

الخطوة الخامسة: اتخاذ القرار:

عند مقارنة الدرجات الحرجة مع درجات العينة من خلال رسمها على نفس توزيعات t وملاحظة ما إذا كانت درجة العينة تقع داخل المنطقة الحرجة أم لا.



شكل رقم (12 ـ 2) الدرجات الحرجة ودرجات العينة

من هنا وبشكل واضح نجد أن الهيئة الوطنية للمعلومات لا تستطيع القول بأن سكان هذه المنطقة لديهم متوسط عمري يصل إلى 40 سنة فقط. بل إن المتوسط العمري

يصل إلى أكثر من 40 سنة ويكون ذا دلالة. ومع أنه بالإمكان سحب عينة بمتوسط عمري 43 سنة من مجتمع متوسطه 40 سنة فقط وهذا عادة ما يحدث فقط في أقل من 5 في كل مئة مرة. آخذين في الاعتبار هذه الاحتمالية المنخفضة بأن التباين العشوائي هو المسبب لوجود متوسط عمري عال للعينة. وبهذه النتيجة فإن الهيئة الوطنية للمعلومات مضطرة لزيادة المخصصات المالية لهذه المنطقة، بالرغم من أن قرارها قد استند على العينة بدلاً من المسح الكلي للسكان في هذا الإقليم. وبالرجوع إلى المثال السابق، يمكننا القول بأن هناك تغيرات طفيفة في إجراء اختبار الفرض الذي تعلمناه في الفصل السابق. إن هذه التغيرات تأخذ في حسابها الحقيقة، أن الانحراف المعياري للمجتمع غير معروف الذي سنقوم بالاستدلال حوله. وفي هذا الخصوص:

- يمكننا استخدام معادلة مختلفة في الخطوة الثالثة لاستنتاج درجة العينة.
- ويمكننا الإشارة إلى فرق بسيط في توزيع المعاينة في الخطوة الرابعة لاستنتاج الدرجة الحرجة والمنطقة الحرجة.

وبمعزل عن تعديلات هذه الإجراءات فإنها أساساً واحدة ولكي نكون على حسن اطلاع أبعد مع اختبار t لمتوسط سوف نقوم بالعمل على الأمثلة التالية:

مثال(1): اختبار أحادي الجانب[3]:

قد تتهم مؤسسة اقتصادية ما بالتعصب حول النساء في سلم التوظيف، وأن أحد الطرق لتقييم ما إذا كان هذا التعصب ضد النساء في الترقي الوظيفي هو النظر إلى متوسط عدد النساء في المواقع الإدارية في كل فروع المؤسسة.

تجدر الإشارة إلى أن لدينا المعرفة من خلال الأرقام المتعلقة بالتشغيل على المستوى الوطني بأن كل المؤسسات المشابهة لديها متوسط 2.5 من النساء في المواقع الإدارية العليا. ومن خلال مسح لعدد 31 من فروع هذه المؤسسة الاقتصادية قد سجل بأن معدل النساء اللائي يشغلن مواقع إدارية يصل إلى 1.9 بانحراف معياري يصل إلى 1. هل نستطيع القول، بأن هذه العينة قد جاءت من مجتمع متوسط توظيف النساء في الإدارة فيه يصل إلى 2.5؟

الخطوة الأولى: صياغة الفرض الصفري والفرض البديل:

Ho: متوسط عدد النساء اللائي يشغلن وظيفة مدير في هذه المؤسسة الاقتصادية هو 2.5

$$Ho: u = 2.5$$

Hi: متوسط عدد النساء اللائي يشغلن وظيفة مدير في هذه المؤسسة الاقتصادية هو أقل من المتوسط في الفروع الأخرى (المؤسسة ضد النساء في عملية التوظيف).

$$Hi: u < 2.5$$

الخطوة الثانية:

في هذا المثال يرغب الباحث في معرفة المتوسط للبيانات المقاسة على المستوى ذي المسافات والنسبي عندما يكون الانحراف المعياري للمجتمع غير معروف، وعليه فإننا نستخدم اختبار t لمتوسط عينة واحدة بدرجة حرية 30.

$$DF = N - 1 = 30$$

الخطوة الثالثة: حساب درجة العينة:

يمكننا تلخيص البيانات المتعلقة بالعينة بالشكل التالي:

$$\overline{X} = 1.9$$

$$N = 31$$

$$S = 1$$

وباستخدام المعادلة لتحويل قيمة العينة إلى درجة t نتحصل على:

$$t = \frac{\overline{x} - u}{\frac{S}{\sqrt{N-1}}} = \frac{1.9 - 2.5}{\frac{1}{\sqrt{31}}}$$

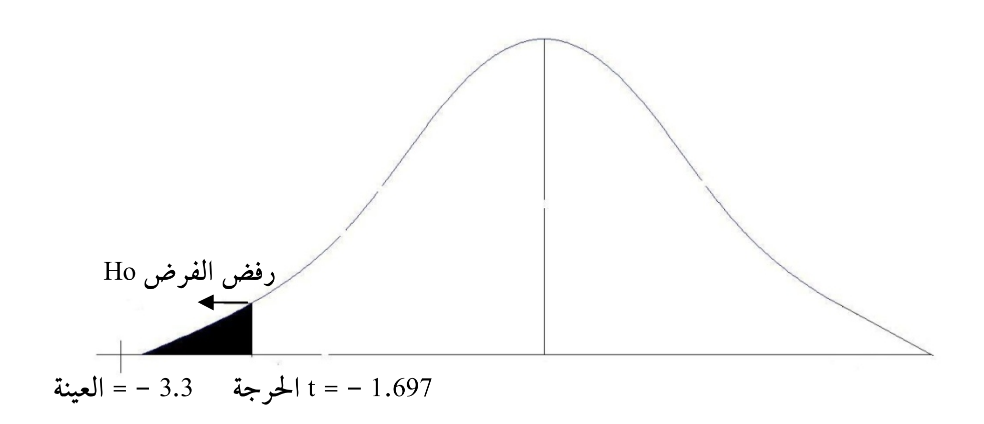
$$= -3.3$$

الخطوة الرابعة: تعيين الدرجة الحرجة والمنطقة الحرجة:

في هذا المثال تم استخدام $\alpha = 0.05$ ولما كنا نرغب في معرفة ما إذا كانت هذه المؤسسة الاقتصادية تتعصب ضد النساء في الترقي الوظيفي. كما أننا أيضاً نستخدم اختبار أحادي الجانب. وبالرجوع إلى جدول توزيع القيم الحرجة لتوزيع t بدرجة حرية 30 فإن القيمة لمستوى الدلالة 0.05 تساوي 1.697-.

الخطوة الخامسة: اتخاذ القرار:

إنه من الواضح بأن قيمة t للعينة تكون قريبة من المتوسط أكثر من القيمة الحرجة كما هو مبين في الشكل التالي:



شكل رقم (12-3) الدرجات الحرجة ودرجات العينة

وعليه، لا يمكننا رفض الفرض الصفري برغم أن متوسط العينة كان أكبر من القيمة المفترضة، ويرجع عدم رفضنا للفرض الصفري إلى احتمالية أن هذا الفرق مرده إلى التباين العشوائي عند عملية المعاينة.

مثال(2): اختبار ثنائي الجانب [4]:

في دراسة متعلقة بأطفال ما بين 5 سنوات و12 سنة في مجتمع ما (أ) تبين أن معدل مشاهدة هؤلاء الأطفال للإذاعة المرئية يصل إلى 196 دقيقة في اليوم. ولغرض التفسير فإننا نفترض أن متوسط ما يشاهده هؤلاء الأطفال هو مقدار ما يشاهده أطفال المجتمع ككل في نفس الفئة العمرية. لقد تم إجراء مسح لعينة عشوائية لمجتمع آخر (ب) تقدر بـ 20 طفلاً من نفس الفئة العمرية لمعرفة ما إذا كان هناك فرق دال بين هؤلاء الأطفال في المجتمعين أ و ب وفيما يتعلق بكمية متوسط ما يشاهده هؤلاء الأطفال في ليلة واحدة.

الفرضية الصفرية:

مفادها أن الأطفال في مجتمع (ب) يشاهدون نفس المعدل الذي يشاهده أطفال مجتمع (أ).

$$Ho: u = 196$$ دقيقة

الفرضية البديلة:

إن الأطفال في مجتمع (ب) يختلف معدل مشاهدتهم للإذاعة المرئية إذا ما قورنوا بأقرانهم في مجتمع (أ).

$$Hi: u \neq 196$$ دقيقة

لاحظ أننا من خلال السؤال البحثي المطروح نرغب في معرفة ما إذا كان هناك فرق دال بين معدل ما يشاهده هؤلاء الأطفال في اليوم في مجتمع (أ) والقيمة المفترضة 196 دقيقة فإننا في واقع الأمر لا ينصب اهتمامنا على ما إذا كان أطفال مجتمع (أ) يشاهدون الإذاعة المرئية بشكل دال أكبر أو أصغر بمجرد أنهم مختلفون. إن الإحصاء الوصفي الذي يلخص لنا نتائج هذا المسح يظهر كالتالي:

$$\overline{X} = 166$$ دقيقة

$$S = 29$$ دقيقة

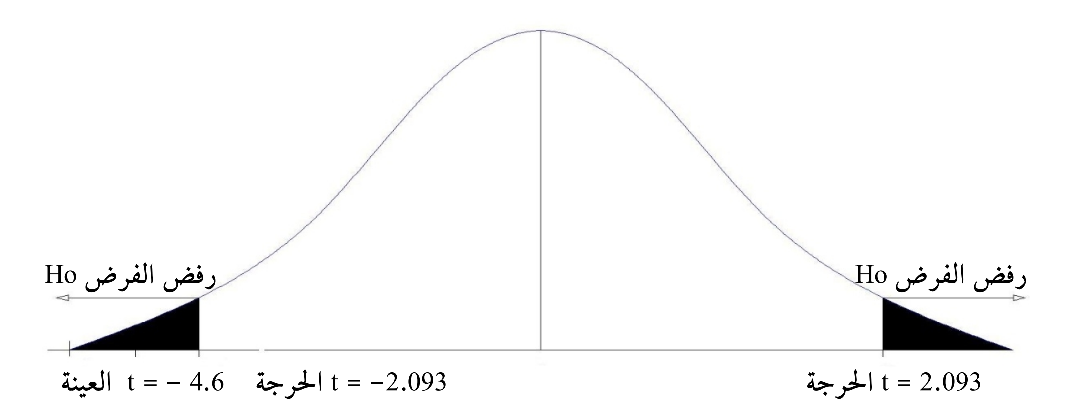
$$N = 20$$

بالتعويض وفقاً للمعادلة التالية نحصل على:

$$t = \frac{\overline{x} - u}{\frac{s}{\sqrt{N-1}}} = \frac{166 - 196}{\frac{29}{\sqrt{20-1}}}$$

$$= -4.6$$

وبالنظر إلى جدول توزيعات t بدرجة حرية 19 على مستوى دلالة 0.05 فإن درجة t الحرجة تساوي 2.093 ± وطالما نحن بصدد اختبار ثنائي الجانب على أساس الفرض البديل، فإن القيمة الحرجة ستكون على جانبي التوزيع. انظر الشكل التالي:



شكل رقم (12ـ4) الدرجات الحرجة ودرجات العينة

من خلال هذا الشكل يمكننا القول بأن هناك فرقاً دالاً بين معدل مشاهدة الأطفال للإذاعة المرئية في ليلة واحدة بين المجتمعين أ و ب.

إجراء توليد اختبار t لعينة واحدة باستخدام Spss:

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Analyze ⟸ Compare Means ⟸ One Sample test

2- اختر TV Watching Per Night من قائمة المتغيرات Variables List.

3- انقر ◄ يقوم بلصق TV Watched Per Night في القائمة المحددة المعنونة بـِ Test Variable (S).

4- في الصندوق المقابل لِـ Test Value: تطبع 196.

5- انقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

T. Test — One - Sample Statistics

	N	Mean	Std. deviation	Std. Error Mean
Tv watched Per Night in Minutes	20	185.85	29.29	6.55

One - Sample test

	Test Value = 196					
	T	Df	Sig (2-tailed)	Mean difference	% Confidence95 Interval of difference	
					Lower	upper
Tv watched Per Night in Minutes	−4.603	19	.000	−30.15	−43.86	−16.44

المصدر: .George Argyrous, op.cit, P. 299

يقدم لنا المربع الأول من اختبار t لعينة واحدة الإحصاءات الوصفية: عدد الحالات (N) 20، متوسط العينة (165.85)، والانحراف المعياري للعينة (29.29). إن آخر رقم في المربع الأول والذي يشير إلى الخطأ المعياري للمتوسط هو الانحراف المعياري لتوزيع المعاينة لـِ t لهذا العدد من درجات الحرية. وهو القيمة في مقام معادلة t.

أما المربع الثاني لاختبار t لعينة واحدة فيحتوي نتائج اختبار الاستدلال حيث إن قيمة (4.603−)، بدرجة حرية df = 19، وبدرجة دلالة لاختبار ثنائي الجانب 0005. (يقربها برنامج Spss إلى ثلاثة أماكن عشرية)، إن الفرق بين قيمة الاختبار لـِ 196، ومتوسط العينة يكون فرق المتوسط 30.15− وهو القيمة في بسط معادلة t.

ولما كانت الاحتمالية منخفضة جداً للحصول على متوسط عينة يقدر بـِ 165.85 أو أقل من مجتمع يشاهد الإذاعة المرئية بمعدل 196 دقيقة في اليوم، فإننا بهذا نرفض الفرض الصفري الذي مفاده أن متوسط المجتمع هو 196 دقيقة لأطفال أستراليا الذين لا يشاهدون الإذاعة المرئية بنفس الكمية التي يشاهدها أطفال بريطانيا.

أما العمود الأخير في المربع الثاني فيشير إلى 95 % فترة ثقة للفرق بين العينة ومدى قيمة الاختبار من الحدود الدنيا 43.86− إلى الحدود العليا 16.44−. بمعنى آخر، إن الفرق بين معدل كمية مشاهدة الإذاعة من مجموعة الأطفال، تقع في مكان ما بين هذا المدى عند مستوى ثقة 95 %. ولما كان هذا المدى لا يشمل قيمة صفر، التي تشير إلى عدم وجود فرق، فإننا بالتالي نستطيع رفض الفرض الصفري الذي مفاده: لا فرق بين المجموعتين في كمية مشاهدة الإذاعة المرئية[5].

الخلاصة:

قد بينا خلال هذا الفصل اختبار t لمتوسط حسابي لعينة واحدة ويشبه هذا الاختبار إلى حد كبير اختبار Z لمتوسط عينة واحدة والذي تم تناوله في الفصل السابق، إلا أنه يمكننا القول بأن هذا الاختبار قد تم في مواقف مألوفة يكون فيها الانحراف المعياري للمجتمع غير معروف. وفي حقيقة الأمر فإن كلا الاختبارين يكونان متطابقين عندما يكون حجم العينة أكبر من 120، وبنظرة للجدول المتعلق بالقيم الحرجة لتوزيع t نجد

على سبيل المثال أن اختبار t ثنائي الجانب على مستوى دلالة 0.05 يساوي 1.96 وهي نفس القيمة لـِ Z على نفس الاختبار وبنفس مستوى الدلالة. بمعنى آخر، عندما يكون حجم العينة أكبر من 120 فإن توزيع t وتوزيع Z يكونان متطابقين تماماً. وعليه فإن المساحات تحت المنحنيات المعينة لأي درجات معطاة تكون أيضاً متطابقة[6].

أسئلة للمراجعة:

1- ما هو الافتراض حول توزيع مجتمع يشكل الأساس لاختبار t؟

2- احسب اختبار t بألفا 0.05 = α لمجموعة البيانات التالية:

	متوسط العينة $\overline{X}$	الانحراف المعياري S	H0	Hi	العينة (N)
a)	62.4	14.1	u = 68	68≠ u	61
b)	62.4	14.1	u = 68	68< u	61
c)	2.3	1.8	u = 3.1	3.1≠ u	25
d)	2.3	1.8	u = 3.1	3.1≠ u	190
e)	102	45	u = 98	98≠ u	210
f)	102	45	u = 90	90≠ u	210

3- من البيانات التالية المتعلقة بالأعمار بالسنوات عند الوفاة لعينة من الناس الذين ولدوا في نفس العام.

, 60 , 72 , 55 , 68 , 12 , 48 , 69 , 78 , 42 , 60 , 81 , 72 , 58 , 70 , 54 , 85 , 34
68 , 74 , 59 , 67 , 76 , 55 , 87 , 70.

- احسب المتوسط العمري عند الوفاة، والانحراف المعياري لهذه البيانات.
- ما هي الاحتمالية العشوائية لحصولنا على عينة من المجتمع بمتوسط الحياة المتوقع لِـ 70 سنة.

4- من البيانات التالية:

.0 , 250 , 300 , 360 , 375 , 400 , 400 , 400 , 420 , 425 , 450 , 462 , 470 , 0
.475 , 502 , 520 , 560 , 700 , 1020

إن المتوسط الحسابي ($\overline{X}$) لهذه البيانات يصل إلى $\overline{X}$ = 424.45 بانحراف معياري يصل إلى S = 216.

المطلوب: إجراء اختبار t بدلالة 05. لتقييم الاحتمالية التي مفادها أن العينة قد سحبت من مجتمع بمتوسط دخل أسبوعياً يصل إلى 480.

5- من جدول توزيعات قيم t أكمل البيانات التالية:

درجة الحرية (df)	الاختبار (test)	الاحتمالية (P)	درجة (t)
5	اختبار أحادي الجانب	...	2.015
10	اختبار ثنائي الجانب	0.02	...
...	اختبار أحادي الجانب	0.05	1.708
65	اختبار ثنائي الجانب	0.05	...
228	اختبار أحادي الجانب	0.10	...

الهوامش والمصادر:

أولا: الهوامش:

1- George Argyrous , Statistics for Social and Health Research , With a Guide to Spss , Sage Publications ,London , 2001 , P. 288

2- Ibid , P. 289.

3- Ibid , P. 294.

4- Ibid , P. 297.

5- Ibid , P. 299.

6- Ibid , P. 300.

ثانيا: المصادر:

1- Frederick J. Gravetter and Larry B. Wallnau , Statistics for The Behavioral Sciencies , 8th ed , USA , 2010.

2- George Argyrous , Statistics for Social and Health Research , With a Guide to Spss , Sage Publications ,London , 2001.

3- Joseph F. Healey , The Essentials of Statistics: ATOOL for Social Research , Wadsworth Cengage Learning , UAS , 2010.

الجزء الرابع

الإحصاءات الاستدلالية البارامترية لعينتين مستقلتين أو أكثر

- الفصل الثالث عشر: اختبار t لعينتين ذات وسطين حسابيين متساويين.
- الفصل الرابع عشر: فروق الدلالة: اختبار **F** لأكثر من وسطين متساويين: تحليل التباين

الفصل الثالث عشر

اختبار t لعينتين ذات وسطين حسابيين متساويين

مقدمة:

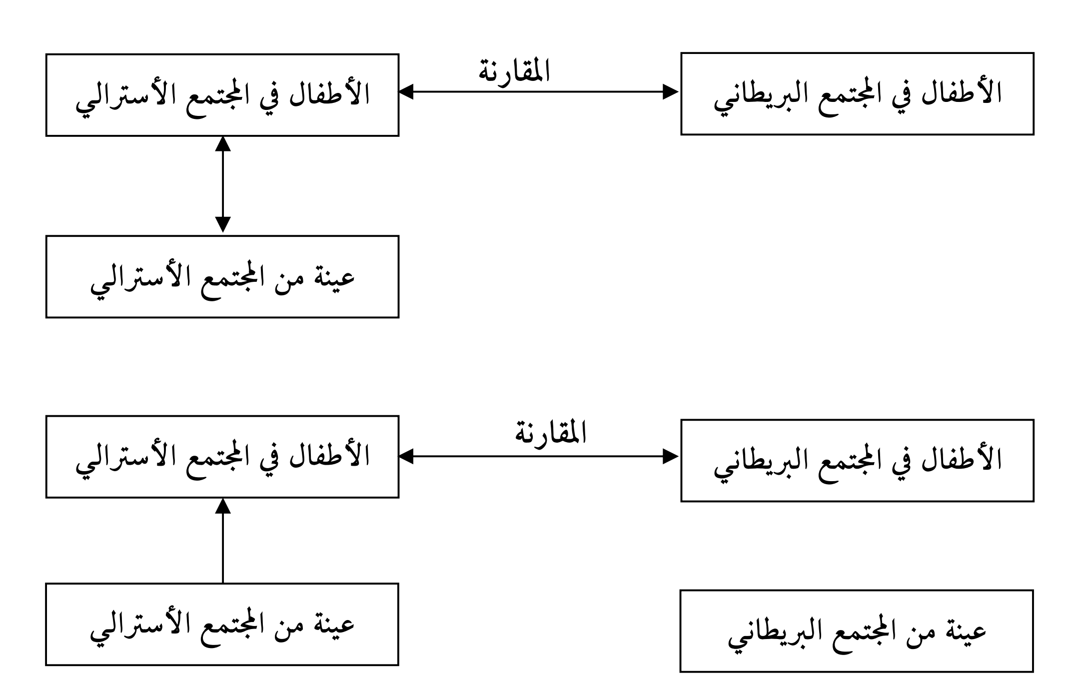
إن ما تناولناه في الفصول السابقة تم التركيز فيه بالدرجة الأولى على الاختبارات المتعلقة بالمتوسط الحسابي لعينة واحدة. بمعنى أن الاهتمام كان يتعلق بالاستدلال حول مجتمع واحد فقط، عندما لم يكن لدينا أية معلومات حول المجتمع، حينئذٍ يمكن استنتاج هذه المعلومات من خلال النتيجة المتحصل عليها من العينة.

في هذا الفصل سيتم التركيز على اختبار الفرض المتعلق بعينتين، ففي هذا الاختبار عادة ما يلجأ الباحث إلى مقارنة مجتمعين فيما يتعلق ببعض الإحصاء الوصفي مثل: المتوسط الحسابي، والنسب ذات الحدين أو التوزيع التكراري.

ولما كانت القيمة لهذه الإحصاءات غير معلومة لكلا المجتمعين، حينئذٍ سيتم سحب عينة من كل مجتمع لغرض الاستدلال من تلك العينات. من ناحية أخرى، إذا كنا لا نعرف القيمة لهذه الإحصاءات لأي من المجتمعين حينئذٍ يمكننا سحب عينة من كل مجتمع وإجراء عملية الاستدلال من كل واحدة من هذه العينات. على سبيل المثال، إذا رجعنا إلى المثال الذي تم طرحه في الفصل المتعلق باختبار t لمتوسط عينة واحدة حيث تم التركيز

حول كمية معدل مشاهدة الإذاعة المرئية (التلفزيون) من قِبل الأطفال في المجتمع الأسترالي والبريطاني للفئة العمرية من 5 إلى 12 سنة. لقد تم التركيز في هذا المثال على مقارنة متوسطات هذين المجتمعين؛ إلا أنه قد لوحظ، أن لدينا المعرفة فقط بالمتوسط المتعلق بالمجتمع البريطاني. ولما كانت القيمة الإحصائية المتعلقة بالمجتمع الأسترالي غير معروفة، الأمر الذي جعلنا نقوم بسحب عينة من 20 مفردة لأجل إجراء عملية الاستدلال استناداً على البيانات التي تحصلنا عليها من هذه العينة (انظر الشكل 13 – 1).

إن السؤال الذي يمكن طرحه هنا، ماذا سيكون الأمر عندما لا تكون لدينا معلومات حول أطفال المجتمع البريطاني. إن أفضل طريقة في هذه الحالة هي سحب عينة عشوائية من هذا المجتمع لإجراء استنتاج آخر من العينة الثانية.



المصدر: George Argyrous, Statistics for Social & Health Research, with a Guide to SPSS , SAGE Publications, London, 2001 , P. 340.

شكل رقم (13 ـ 1) اختبار الفرض لحالة عينتين

في مثل هذا الموقف نحن نقوم بإجراء اختبار الدلالة لعينتين. في هذا المثال، نحن نقوم بمسح الأطفال في كل واحد من هذين المجتمعين، بالرغم من أنه من الناحية العملية نحن نفكر في واقع الأمر في عينة واحدة تتعلق بالمجتمعين الأسترالي والبريطاني. وتصورياً، نحن نقول إننا نتعامل مع عينتين من كل مجتمع من هذين المجتمعين للمقارنة بينهما. وبالرغم من الآلية الواقعية لجمع البيانات، فإننا نقوم بجمع بيانات ضخمة من الأطفال الذين تم مسحهم كجزء من عملية البحث ذاتها، إلا أنه عند تحليل البيانات، فإننا في واقع الأمر نتعامل مع مجموعتين من الأطفال كعينتين منفصلتين، كما هو واضح في الشكل أعلاه.

في حقيقة الأمر، يمكننا توسيع هذا الموقف إذا أردنا أن نقارن أكثر من مجتمعين. على سبيل المثال، قد نكون راغبين في مقارنة معدل مشاهدة الأطفال للإذاعة المرئية في أكثر من مجتمعين فيما يتعلق بمعدل مشاهدتهم للإذاعة المرئية، ولدينا فقط عينات من كل واحد من هذه المجتمعات.

إن العمل مع أكثر من عينتين يتطلب إجراء اختبار مختلف ذي دلالة وهذا ما سنتناوله في الفصل اللاحق من هذا الكتاب.

بصفة عامة، إن اختبار الاستنتاج يتأثر بعدد العينات التي يتم الاستدلال منها. وجدير بالذكر أنه يمكننا القول، بأنها ممارسة شائعة أن نميز بين اختبارات عينة واحدة، أو اختبارات عينتين أو اختبارات أكثر من عينتين. وعند الحديث عن الاستدلال من أكثر من عينتين، فإننا بذلك نتحدث عن K من العينات K-Samples، حيث تشير K إلى العدد الذي يزيد عن اثنين. في أحوال كثيرة فإن التغير المرتبط في الانتقال من موقف اختبار t لعينة واحدة إلى اختبار t لعينتين أو لموقف K من العينات لن تكون تغيرات كبيرة، بقدر ما تكون عملية أساسية لتنظيم، وعليه سيكون من المفيد لنا معرفة ما إذا كان عدد العينات التي على ضوئها تمت عملية الاستدلال، هل هي عينة واحدة أم عينتين أم أكثر من ذلك.

دعنا مرة ثانية نرجع إلى المثال المتعلق بمعدل مشاهدة الأطفال للإذاعة المرئية

(التلفزيون) للمقارنة بين أطفال المجتمع الأسترالي وأطفال المجتمع البريطاني. إننا كنتيجة لذلك، نقوم بجمع بيانات حول متغيرين: مكان الإقامة، ومعدل كمية مشاهدة الإذاعة المرئية. بمعنى آخر، هل يمكن أن يختلف الطفل عن طفل آخر في واحدة من الطريقتين؟ فالطفل يمكن أن يختلف فيما يتعلق بمكان الإقامة و / أو يمكن أن يختلف فيما يتعلق بكمية ما يشاهده[1].

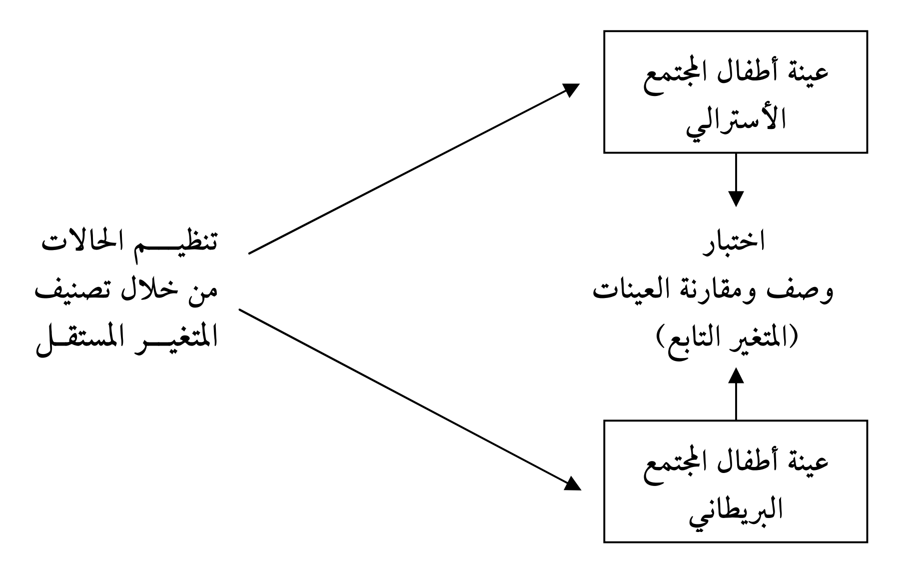
المتغيرات المستقلة والمتغيرات التابعة

Dependent and Independent Variables

يمكننا أن نفكر في مشكلة هاتين العينتين طبقاً للمعطيات المتعلقة بالمتغيرات المستقلة والمتغيرات التابعة والتي أشرنا إليها في متن هذا الكتاب (الفصل الأول). عادة ما يطلق على المتغير التصنيفي المتغير المستقل، ومتغير الاختبار يطلق عليه المتغير التابع. فالمتغير التابع هو ذلك المتغير الذي يفسر أو يتأثر بالمتغير المستقل.

وبالرجوع مرة أخرى للمثال المتعلق بالأطفال وكمية مشاهدة الإذاعة المرئية في كل ليلة، فإننا نشك في أن مكان الإقامة بطريقة ما يؤثر أو يسبب تأثيراً في كمية مشاهدة الأطفال للإذاعة المرئية (يمكن أن تكون هناك عوامل أخرى مثل الطقس، أو نوعية البرامج المقدمة في بلدان مختلفة). إنه من الواضح أنه في هذا الموقف يوجد لدينا حالة سببية في اتجاه واحد one way Causality يجب أن تنطلق من مكان الإقامة إلى مشاهدة الإذاعة المرئية، إنه من غير المعقول أن عادة مشاهدة الأطفال للإذاعة المرئية تُحَدَّدُ من خلال المكان الذي يعيشون فيه، في شواهد أخرى، على أية حال، إن اختبار النموذج الملائم يمكن أن يكون أكثر جدلاً كما بيّنّا في الفصل السادس).

تجدر الإشارة إلى أن كل هذه الاعتبارات مرتبطة بتنظيم البيانات في حالة عينتين كما يلخصها الشكل التالي:



المصدر: George Argyrous, Statistics for Social & Health Research, op.cit., P. 341.

شكل (13 ـ 2) اختبار الدلالة لعينتين

توزيعات المعاينة للفرق بين وسطين:

كما هو الحال في كل الاختبارات المتعلقة بالفروض، فإننا نبدأ بافتراض أن الفرض الصفري Ho مفاده أنه لا يوجد فرق أنه فرض صحيح. وعلى ضوء هذه الفرضية يمكننا بناء توزيع المعاينة لنحدد احتمالية الحصول على الفرق المشاهد بين متوسطي عينتين من مجتمعات من دون فرق أي متساوية في متوسطاتها.

وأخيراً، نقوم بمقارنة هذه الاحتمالية مع مستوى ألفا (α) لنقرر ما إذا كنا قادرين على رفض الفرض الصفري.

مثال: دعنا نبدأ بافتراض أن معدل مشاهدة الأطفال للإذاعة المرئية متساوٍ في كلا المجتمعين - الأسترالي والبريطاني - إن هذا الفرض الصفري الذي مفاده لا يوجد فرق يمكن التعبير عنه جبرياً كالتالي:

$$H_0:\ \mu_1 = \mu_2$$

أو $$H_0:\ \mu_1 - \mu_2 = 0$$

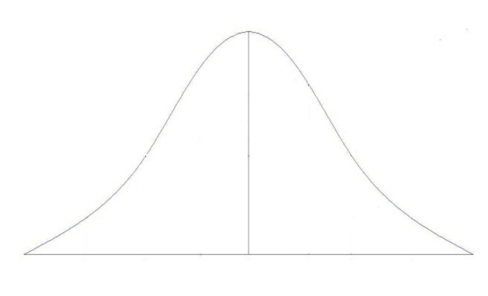
إذا كانت هذه الفرضية فرضية صحيحة، ما الذي نحصل عليه إذا سحبنا عينات مكررة من كل مجتمع وقمنا بحساب الفرق في المتوسطات لكل زوج من العينات؟

إدراكياً، نحن نتوقع أن النتيجة الأكثر شيوعاً بأن الفرق سيكون صغيراً، أنه لم يكن صفراً (0). وبما أننا افترضنا أنه لا يوجد فرق بين وسطي المجتمعين، فإننا بالتالي نتوقع أن وسطي العينتين سيكونان متساويين أيضاً.

$$\bar{X}_1 = \bar{X}_2 \quad \therefore \bar{X}_1 - \bar{X}_2 = 0$$

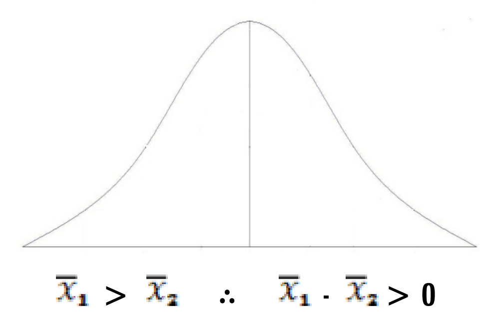
انظر الشكل التوضيحي التالي:

عينة المجتمع الأسترالي



$$\bar{X}_1 = \bar{X}_2 \quad \therefore \quad \bar{X}_1 - \bar{X}_2 = 0$$

عينة المجتمع البريطاني

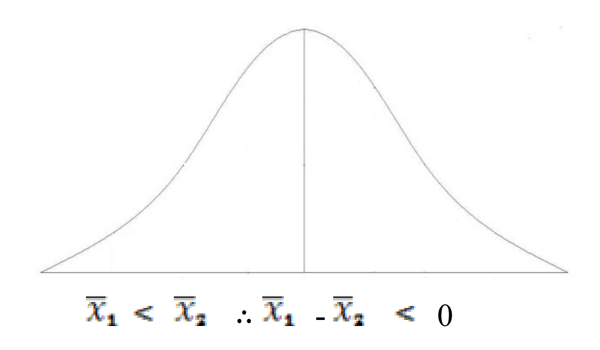


$$\bar{X}_1 > \bar{X}_2 \quad \therefore \quad \bar{X}_1 - \bar{X}_2 > 0$$

غير أنه يمكننا القول في هذا السياق، أنه ليس في كل الأحوال أن تكون النتيجة دائماً هكذا. أحياناً، يمكننا سحب عينة من المجتمع الأسترالي بمعدل أقل لما يشاهده الأطفال في المجتمع البريطاني، كما يوضحه الشكل التالي:

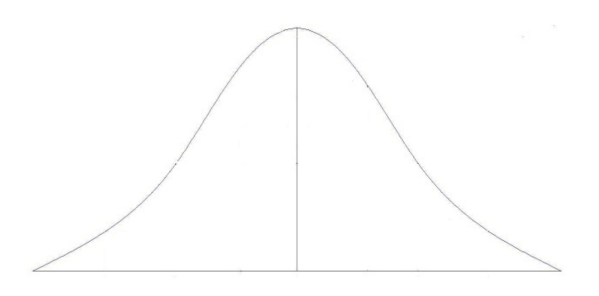
(عينتان بمتوسطين غير متساويين).

عينة المجتمع الأسترالي



$$\bar{X}_1 < \bar{X}_2 \quad \therefore \bar{X}_1 - \bar{X}_2 < 0$$

عينة المجتمع البريطاني



وبنفس الطريقة، يمكننا الحصول من خلال عملية الفرصة العشوائية على موقف مختلف:

$$\bar{X}_1 > \bar{X}_2 \quad \therefore \bar{X}_1 - \bar{X}_2 > 0$$

إذا أخذنا عدداً كبيراً من العينات المتكررة، وقمنا بحساب الفرق بين كل زوج من متوسطات العينة، فإننا سننتهي بتوزيع المعاينة للفرق بين وسطي عينتين لديها الخصائص التالية:

- سيكون توزيع t: $t = \frac{\bar{x}_1 - \bar{x}_2}{\sigma \bar{x} - \bar{x}}$
- أن فرق المتوسط بين وسطي عينتين سيكون صفراً (0)

$$\mu\bar{x} - \bar{x} = 0$$

- أن انتشار الدرجات حول هذا المتوسط الذي قيمته صفر سيعرف بالمعادلة التالية:

$$\sigma\bar{x} - \bar{x} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

ويطلق على هذه العملية التقدير المجمع للتباين Pooled Variance Estimate ويفترض هذا التقدير أن المجتمع لديه تباينات متساوية Equal Variances. إلا أن هذه الفرضية، في بعض الحالات، لا يمكن أن تكون ثابتة. في مثل هذه الحالة، يمكن استخدام تقدير تباين منفصل. ومن خلال برامج Spss يمكننا بسهولة حساب قيمة t مستخدمين كلا التقديرين، إضافة إلى المعلومات التي تمكننا من اختبار إما تباينات متساوية أو تباينات غير متساوية. ولكن عندما نقوم بحساب قيمة t يدوياً فإن التقدير المجمع للتباين، بصفة عامة يستخدم لأنه أكثر سهولة للعمل به، وسوف يقودنا في العادة إلى نفس القرار الذي وَصَلْنَا إليه في حالة التباين المستقل[2].

اختبار t لعينتين مستقلتين ذات وسطين متساويين:

تجدر الإشارة هنا إلى أنه يمكننا استخدام خاصيات توزيع المعاينة لإجراء اختبار t لمتوسطات حسابية متساوية.

مثال: نفترض أنه لدينا مسح يحتوي على 20 طفلاً من المجتمع الأسترالي. وعشرون طفلاً من المجتمع البريطاني، وأننا نرغب في تقييم ما إذا كان وقت مشاهدة

الإذاعة المرئية يتأثر بمكان الإقامة (لاحظ أنه ليس بالضرورة أن يكون عدد الحالات متساوية في العينتين).

دعنا الآن نجري اختبار t وفقاً للخطوات الخمس المرتبطة باختبار الفرض الإحصائي:

1- الخطوة الأولى: صياغة الفرض الصفري والفرض البديل.

الفرض الصفري: لا يوجد فرق في معدل مشاهدة الأطفال للإذاعة المرئية في المجتمعين الأسترالي والبريطاني.

$$H_0 : \mu_1 = \mu_2 \quad \text{أو} \quad H_0 : \mu_1 - \mu_2 = 0$$

الفرض البديل: يوجد فرق في معدل مشاهدة الأطفال للإذاعة المرئية في المجتمعين الأسترالي والبريطاني.

$$H_1 : \mu_1 - \mu_2 \neq 0 \quad \text{أو} \quad H_i : \mu_1 \neq \mu_2$$

2- الخطوة الثانية: اختيار اختبار الدلالة:

هناك عاملان أساسيان يمكن أخذهما في الاعتبار عند اختيار اختبار الدلالة:

أ- أننا نقوم بعملية الاستدلال من عينتين: عينة أطفال أستراليا وعينة أطفال بريطانيا. ومن هنا ينبغي علينا إجراء اختبار t لعينتين.

ب- ينبغي مقارنة العينتين في ضوء معدل الوقت الذي يقضيه الطفل في مشاهدة الإذاعة المرئية. وقد تم قياس هذا المتغير على المستويين ذي المسافات المتساوية والنسبي. ومن هنا فإن الإحصاءات الملائمة، هي إحصاءات وصفية تتمثل في المتوسط الحسابي لكل من العينتين.

إن هذين العاملين يقودانا إلى اختيار اختبار t لعينتين ذات وسطين حسابيين متساويين كاختبار ملائم للدلالة.

3- الخطوة الثالثة: حساب درجات العينة:

إن النتائج التالية تصف البيانات لكل عينة (حيث تمثل العينة رقم 1 المجتمع الأسترالي، والعينة 2 المجتمع البريطاني):

$\overline{X}_2 = 187$ دقيقة	$\overline{X}_1 = 166$ دقيقة
$S_2 = 30$	$S_1 = 29$ دقيقة
$N_2 = 20$	$N_1 = 20$

ووفقاً للمعادلة التالية نتحصل على درجة t:

$$t_{\text{العينة}} = \frac{\overline{x}_1 - \overline{x}_2}{\sigma \overline{x} - \overline{x}}$$

حيث أن:

$$\sigma\overline{\mathbf{x}} - \overline{\mathbf{x}} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

بالتعويض نتحصل على:

$$\sigma\overline{\mathbf{x}} - \overline{\mathbf{x}} = \sqrt{\frac{(20 - 1)29^2 + (20 - 1)30^2}{20 + 20 - 2}} \sqrt{\frac{20 + 20}{(20)(20)}}$$

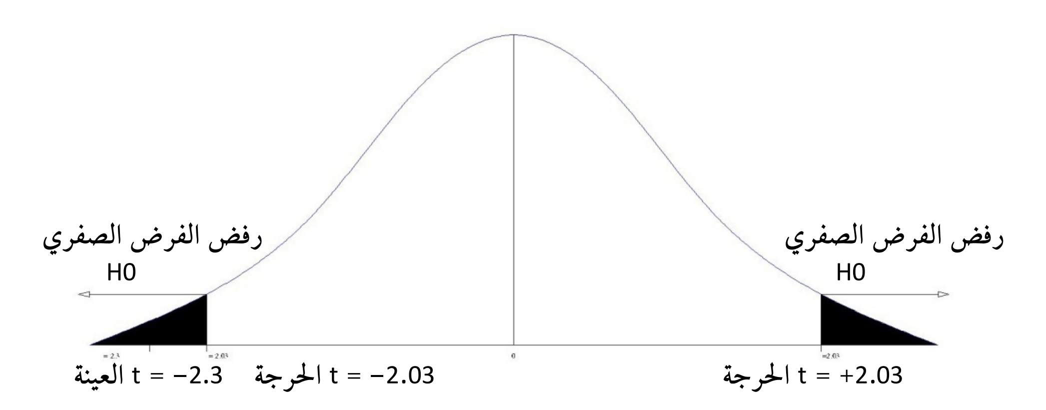
$$= 9.3$$

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sigma \overline{x} - \overline{x}} = t_{\text{العينة}} = \frac{166 - 187}{9.3} = -2.3$$

4- الخطوة الرابعة: اختيار الدرجات الحرجة والمنطقة الحرجة:

بما أننا قد افترضنا أن تباينات العينة مساوية لتباينات مجتمع غير معروفة. وهنا قد افترضنا قيدين على البيانات. وهذا يعني أن درجة الحرية df = n-2 = 40 - 2 = 38

ومن خلال البحث في جدول توزيع t بدرجة حرية 38 على مستوى الفا (α) 0.05، وباستخدام اختبار ثنائي الجانب (استناداً على الفرض البديل) فإن القيمة الحرجة لِـ t تساوي 2.03 (مقابل df = 35). لاحظ أن الجدول لا يحتوي على صف باحتمالية 38 درجة حرية. في مثل هذا الموقف نشير إلى الصف القريب من درجة الحرية التي تعنينا وهي في هذا المثال 35 (انظر الجدول 13 – 1).



شكل (13 ـ 3) الدرجة الحرجة ودرجة العينة

جدول (13 ـ 1) القيم الحرجة لتوزيعات t

	مستوى الدلالة لاختبار أحادي الجانب				
	0.10	0.05	0.02	0.01	0.005
	مستوى الدلالة لاختبار ثنائي الجانب				
Df	0.20	0.10	0.05	0.02	0.01
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.56	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.469	2.763
29	1.311	1.699	0.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
35	1.306	1.690	2.030	2.438	2.724
40	1.303	1.684	2.021	2.423	2.704
45	1.301	1.679	2.014	2.412	2.690
50	1.299	1.676	2.009	2.403	2.678
55	1.297	1.673	2.004	2.396	2.668
60	1.296	1.671	2.000	2.390	2.660
70	1.294	1.667	1.994	2.381	2.648
80	1.292	1.664	1.990	2.374	2.639
90	1.291	1.662	1.987	2.368	2.632
120	1.289	1.658	1.980	2.358	2.617
∞	1.282	1.645	1.960	2.326	2.576

5- الخطوة الخامسة: اتخاذ القرار:

لما كانت درجة العينة تقع داخل المنطقة الحرجة، عليه نرفض الفرض الصفري Ho (انظر الشكل رقم 13) وبالتالي لا يمكننا القول بأن أطفال المجتمع الأسترالي وأطفال المجتمع البريطاني لديهم نفس معدل مشاهدة الإذاعة المرئية في الليلة الواحدة [3].

إجراء اختبار t لعينتين مستقلتين باستخدام برنامج Spss :

الإجراء:

1- في القائمة الموجودة في الجزء العلوي من الشاشة انقر فوق:

Anayze ⟸ Compare Means ⟸ Independent Sample

t- test (هذه العملية تعرض لنا مربع الحوار Independent Sample t- test)

2- انقر على المتغير التابع Minutes of TV Watched في القائمة (Source List)

3- انقر على ◄ التي تشير إلى قائمة متغيرات الاختبار (تقود هذه العملية إلى لصق (Pastes) Minutes of TV Watched في قائمة Test Variable(s)

4- انقر فوق Country of Residence في القائمة.

5- انقر على ◄التي تشير إلى قائمة Grouping Variable (هذه العملية تلصق Country of Residence في قائمة Grouping Variable

6- انقر على Define Group

7- المساحة التالية لـِ Group1 يتم طبع 2، والمساحة الثانية لـِ Group2 يتم طبع 3. وهي في هذه الحالة Australia و Britain

8- انقر فوق Continue

9- انقر فوق OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

T. Test

Group Statistics

Country Residence	N	Mean	Std. Deviation	Std. Error Mean
Watch of TV. Australia	20	165.85	29.29	8.55
Watch of TV. Britain	20	189.75	29.56	8.81

Independeat Samples test

Watch tv per night	Leven`s test for equa lity of variance		T	Df	Sig -tailed2	Mean difference	Std. Error difference	confidence %95 Interval of the difference	
	F	Sig						Lower	Upper
equal variances assumed	.025	.876	2.24 -6	38	.031	-20.90	9.31	-39.74	-2.06
Egual variances☐ Not assumed			2.24 -6	37.97	.031	-20.90	9.31	-39.74	-2.06

George Argyrous, Statistics for Social & Health Research, op.cit. , PP. 348-349 **المصدر:**

شكل (13 ـ 4) مخرجات SPSS لاختبار t (Spss t- test out put)

تفسيرات مخرجات اختبار t:

يعطي المربع الأول من مخرجات اختبارات العينات المستقلة Independent Samples test والمعنون بـ Group Statistics الإحصاء الوصفي: عدد الحالات (N) Number of Cases، المتوسط الحسابي The mean، والانحراف المعياري The Standard deviation لكل مجموعة.

ويعطي المربع الثاني من مخرجات اختبار العينات المستقلة Independent Samples test، الإحصاءات الاستدلالية. ويعطي هذا المربع معلومات لاختبارين مختلفين لـِ t: أولهما يشير إلى التحقق من فرضية التباينات المتساوية للمجتمع Equal Variances assumed، وثانيهما التحقق من فرضية التباينات غير المتساوية Equal Variances not assumed.

في هذا المثال الذي تعاملنا معه في هذا الفصل تم افتراض أن التباين في المجتمعين المقارنين متساوٍ.

تجدر الإشارة إلى أنه من الناحية العملية هذا يعني استخدامنا لتقدير التباين المشترك Pooled Variance Estimate للعمليات الحسابية، ومع ذلك قد لا يكون الفرض صحيحاً. إن صدق هذه الفرضية تم اختبارها في العمود المعنون بِـ اختبار Leven`s test for Equality of Variance لتساوي التباينات. إن قيمة F هي نسبة تباين العينتين (المجموعتين) وإذا لم تكن هذه النسبة مساوية لواحد (1)، فإنها يمكن أن تعكس الفرق الضمني في تباينات المجتمع. فإذا كانت قيمة F دالة (عمود .sig) على مستوى 05. أو أقل (على سبيل المثال، 01. و 001.)، يعني ذلك أن فرق التباينات المشاهدة في العينتين يعكس الفرق في تباينات المجتمع التي سُحِبَتْ منه هاتان العينتان. وفي مثل هذا الموقف يمكننا أن نشير إلى درجة t (T. Score) في الصف الأول من المربع. وباختصار، يمكننا اتبّاع الإرشادات التالية:

نقرأ عبر الصف الأول الذي يشير إلى تحقق فرضية التباينات المتساوية Equal Variance assumed، فإذا كانت قيمة أقل من 0.05، (0.01، 0.001، 0.03، 0.04... الخ)، يعني ذلك أن فرضية التباينات غير المتساوية Equal Variances not assumed. أما إذا كانت قيمة sig أكبر من (مثلاً 0.06، 0.07، 0.08، 0.09 و 0.10) يعني تحقق فرضية التباينات المتساوية Equal Variances assumed، وبالتحرك عبر الأعمدة نجد أن قيمة t تساوي 2.246– (T. Value is - 2.246) بدرجة حرية (df) 38 بدلالة 0.031 (ثنائي الجانب) (tailed2 -). ولما كانت احتمالية العينة تساوي 0.031 أقل من مستوى ألفا (α) 05.، عندئذٍ يمكننا رفض الفرض الصفري الذي مفاده أن تباينات المجتمع متساوية.

كذلك لدينا عمود فرق المتوسط Mean difference. وهو الفرق بين متوسط العينتين (20.9–) أي ($\overline{X}_2 - \overline{X}_1$).

كذلك يمكننا أن نلاحظ أن مخرجات Spss تولد 95 % فترة ثقة (95%

(Confidence Interval) لفرق متوسطات العينة وهو: 39.74 - 2.06) هذا يسمح لنا بإجراء نفس اختبار الاستدلال: انظر الفصل العاشر. وتشير 95 % فترة الثقة إلى أنَّ الفرق بين متوسطات المجتمع يقع في مكان ما بين 39.74- دقيقة و2.06- دقيقة.

إن فترة الثقة هذه لا تحتوي قيمة (0)، عندئذٍ يمكننا رفض الفرض الصفري الذي مفاده أن متوسطات المجتمع متساوية [(4)].

حساب حجم التأثير لاختبار العينات المستقلة:

نعني بحجم التأثير هنا Effect Size "بأنه عبارة عن مجموعة من الإحصاءات التي تحدد القوة النسبية للفروق بين قيم الوسط الحسابي. بعبارة أخرى، يصف حجم التأثير مقدار التباين الكلي في المتغير التابع، الذي يمكن التنبؤ به من معرفة مستويات المتغير المستقل"[(5)].

ويعطي إحصاء حجم التأثير مؤشراً لحجم الفروق بين المجموعات (وليس مجرد تحديد ما إذا كان الفارق قد حدث على سبيل الصدفة أم لا). والإحصاء المستخدم لهذا الغرض هو إحصاء Eta^2 إيتا تربيع. ولما كانت مخرجات Spss لا تقوم بحساب Eta^2 في اختبار t، فالباحث يمكنه حسابها من المخرجات المرتبطة باختبارات t وفقاً لإحدى المعادلتين:

$$Eta^2 = \frac{t_2}{t_2(N_1 + n_2 - 2)}$$

أو

$$r^2 = \frac{t_2}{t_2 + df}$$

$$= \frac{5.0445}{5.0445 + 38}$$

$$= \frac{5.0445}{43.0445}$$

$$= 0.117$$

وبناء على ما افترضه Cohen`s d [6] لتفسير هذه القيمة فإن:

0.01 تأثير ضئيل

0.06 تأثير معتدل

0.14 تأثير كبير

وفي هذا المثال، يتبين لنا أن حجم التأثير يبلغ 0.117 وهو تأثير معتدل. أي أن متغير مكان الإقامة يبين نسبة 0.12 في المائة فقط من تباين الفروق في متغير مشاهدة الإذاعة المرئية في المجتمعين الأسترالي والبريطاني.

بالإضافة إلى طريقة كوهينز d لقياس حجم التأثير فقد طوّر أيضاً معياراً آخر لتقييم تأثير حجم المعالجة ويتمثل هذا المعيار في (r^2). وتقترح هذه الطريقة تقييم حجم الارتباط، ولكنها سرعان ما توسعت لتطبق على r^2.

تأثير صغير	0.01	r^2 =
تأثير متوسط	0.09	r^2 =
تأثير كبير	0.25	r^2 =

اختبار أحادي الجانب وثنائي الجانب One - Tail and Two Tail Tests:

في المثال السابق لم يكن لدينا أي سبب لأن نشك في أنَّ أيّاً من أطفال المجتمعين يشاهدون أكثر من الأطفال الآخرين كما يعكسه الفرض البديل الذي تمتْ صياغته بشكل بسيط أنه يوجد فرق.

$$Hi : \mu_1 \neq \mu_2$$

وبناءً على هذه المعلومات المعطاة تم إجراء اختبار ثنائي الجانب.

إلا أنه يمكننا القول بأن الجو الدافئ الذي يسود المجتمع الأسترالي يمكن أن يقلل من رغبة الأطفال في مشاهدة الإذاعة المرئية أكثر مما يفعله أطفال المجتمع البريطاني. من هنا لا

نتوقع الفرق، فقط، وإنما أيضاً نبحث عن اتجاه هذا الفرق. ويقودنا هذا الموقف إلى صياغة الفرض البديل بالشكل التالي:

$$Hi : \mu_1 < \mu_2$$

حيث إن u_1 هي متوسط أطفال المجتمع الأسترالي، و u_2 هي متوسط أطفال المجتمع البريطاني (لاحظ أن هذه الصياغة للفرض البديل تقودنا إلى إجراء اختبار أحادي الجانب).

مثال لزيادة التوضيح:

$N_2 = 50$ $N_1 = 50$

$S_2 = 1.2$ $S_1 = 1.3$

$\bar{X}_2 = 3.5$ $\bar{X}_1 = 4.2$

الخطوة الأولى: صياغة الفرض الصفري والفرض البديل:

الفرض الصفري H_O لا يوجد فرق بين وسطي العينتين

$H_0 : \mu_1 - \mu_2 = 0$ أو $H_0 : \mu_1 = \mu_2$

الفرض البديل H_i عينة 1 تسجل معدل أكبر من عينة 2

$H_i : \mu_1 - \mu_2 > 0$ أو $H_i : \mu_1 > \mu_2$

الخطوة الثانية: اختيار درجة الدلالة.

لما كان اهتمامنا ينصب على الاستدلال من عينتين، فإننا بالتالي نقارن بين وسطين. إن الإحصاء الملائم هو المتوسط الحسابي لكل عينة، وبالتالي فإننا نستخدم اختباراً لعينتين ذات وسطين حسابيين كاختبار ملائم للدلالة.

الخطوة الثالثة: حساب درجة العينة:

إن أول خطوة نحتاجها هي حساب الخطأ المعياري Standard Error مفترضين تباينات مجتمع متساوية Assuming Equal Variances.

$$\sigma_{\bar{x}-\bar{x}} = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}\sqrt{\frac{n_1+n_2}{n_1 n_2}} \quad \text{(الخطأ المعياري)}$$

بالتعويض نتحصل على:

$$\sigma_{\bar{x}-\bar{x}} = \sqrt{\frac{(50-1)1.3^2+(50-1)5_2^2}{50+50-2}}\sqrt{\frac{50+50}{(50)(50)}}$$

$$= .250$$

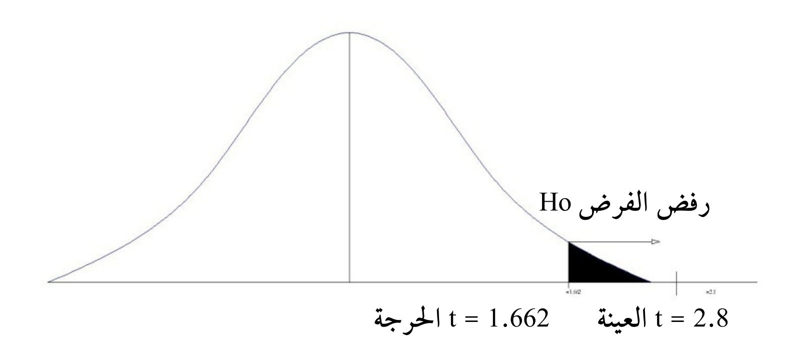
$$t = \frac{\bar{x}_1-\bar{x}_2}{\sigma_{\bar{x}-\bar{x}}} = \frac{4.2-3.5}{0.25} = 2.8$$

الخطوة الرابعة: اختيار درجة الحرية والمنطقة الحرجة:

عندما تجمع العينتان 50 + 50 - 2 = 100 - 2 = 98 درجة حرية وذلك استناداً على الفرضية البديلة، فإننا نستخدم اختباراً أحاديَّ الجانب (الذيل الأيمن). وبالرجوع لجدول توزيع t للقيم الحرجة، بدرجة حرية 98، ومستوى دلالة 0.05 مستخدمين اختباراً أحاديَّ الجانب، فإن القيمة الحرجة تساوي 1.662 + (لاحظ أن الجدول لا يحتوي على صف لاحتمال درجة 98، فإننا في هذا الموقف نشير إلى درجة حرية أقل من الرقم المرغوب وهي 90.

الخطوة الخامسة: اتخاذ القرار:

يمكننا القول أن درجة t للعينة أبعد من صفر إذا ما قورنت بالدرجة الحرجة انظر الشكل رقم (13 – 5). بمعنى آخر، أن درجة t للعينة تقع في منطقة الرفض. وعليه، فإننا نرفض الفرض الصفري الذي مفاده لا يوجد فرق[7].



شكل (13 ـ 5)

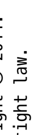
العينات المستقلة والعينات التابعة:

كما أشرنا في الجزء الأول من هذا الفصل بأن العينات المستقلة هي تلك العينات حيث محكات اختيار الحالات التي تشكل العينة الواحدة لا تتأثر بمعايير الاختيار التي تشكل العينة الثانية. بمعنى آخر، أن العينتين مستقلتان عن بعضهما البعض. أي أن تركيبة أحد العينات لا تتساوى مع تركيبة العينة الأخرى. وعليه فإن العيّنتَيْن تعكسان نوعين منفصلين من المجتمع(8). فعلى سبيل المثال، كما بينا سابقاً، يمكننا مقارنة عينة عشوائية للأطفال من المجتمع الأسترالي وعينة من أطفال المجتمع البريطاني، وذلك لاستقصاء علاقة مكان الإقامة بمعدل مشاهدة الإذاعة المرئية.

أما العينات التابعة فهي تلك العينات التي يتم فيها اختيار الحالات التي تكون أفراد العينة الواحدة وفقاً لمعايير تؤثر في اختيار الحالات المكونة للعينة الأخرى، بمعنى، أن أفراد عينة واحدة لم يتم اختيارهم بشكل مستقل، وإنما يتم تحديدهم من خلال تركيبة عينة أخرى(9).

وتجدر الإشارة إلى أن هناك موقفين عامَّين مطلوبَيْن لمثل هذه التبعية:

- عندما يلاحظ نفس الموضوع تحت شرطين مختلفين، وهذا في أحوال كثيرة يتم استخدامه في التجربة (القبلية - البعدية) (before and after experiment) (في



بعض الأحيان يطلق عليه الاختبار القبلي والاختبار البعدي) (a pre - test - post test) على سبيل المثال، يمكن اختبار دواء جديد لمعرفة مدى تأثيره على مرض ضغط الدم. فالمجموعة التي تعاني من ضغط الدم كمجموعة للدراسة تخضع لقياس ضغط الدم، ثم بعد ذلك تعطى جرعة من الدواء الجديد، وبعدها يتم مرة ثانية قياس ضغط الدم. وبجلاء لكي يتم عزل تأثير الدواء، فإن الشخص الذي خضع كعينة للعلاج القبلي، هو نفسه يخضع كعينة للعلاج البعدي. إن القياس لكل شخص في العينة القبلية هو إذاً يتناغم Matched مع خصوصية القياس بعد تناول الدواء الجديد لنرى ما إذا كان هناك تحسن في وضعهما.

- عندما تكون الموضوعات في عينات مختلة مرتبطة بسبب خاص. على سبيل المثال إذا أردنا أن نقارن كمية مشاهدة الإذاعة المرئية للأب، بكمية مشاهدة لإذاعة المرئية لطفله. وإذا اخترنا مجموعة من الآباء، لا نستطيع اختيار أي مجموعة من الأطفال لمقارنتهم بِـ: العينة تحتاج إلى أن تحتوي على الأطفال الذين آباؤهم يشكلون عينة الدراسة. ويطلق على هذا في بعض الأحيان أسلوب العينات المتماثلة Matched - Pairs.

إنه من الواضح، أن أحد المواقف يشكل أحد العينات التي بدورها تحدد شكل العينة الأخرى.

إن ميزة طريقة العينة التابعة هي أنها تتحكم بطريقة غير محكمة في المتغيرات الأخرى التي يمكن أن تؤثر في المتغير التابع. على سبيل المثال، دعنا نفكر أكثر في موضوع ما إذا كان الآباء والأطفال يختلفون في كمية مشاهدة الإذاعة المرئية. فإذا سحبنا عينة عشوائية من الآباء، وأخرى من أي من الأطفال وقارنا بين المتوسطات في كل عينة من هاتين العينتين، فإنه يمكن أن نتحصل على نتيجة مفادها أن هناك فروقاً ذات دلالة إحصائية. إلا أن هذه الفروق قد لا ترجع للوضع العائلي، بقدر ما ترجع ربما إلى متغير آخر كالوضع الاجتماعي والاقتصادي الذي يؤثر بدوره في كمية مشاهدة الإذاعة المرئية. وبما أن عينة الآباء تتألف من حالات كثيرة جاءت من مجموعة اقتصادية واحدة أكثر من عينة الأطفال، وبالتالي ظهر الفرق للعيان.

وتجدر الإشارة إلى أنه يمكن افتراض أن أي زوج معطى من الآباء أو الأطفال يقع داخل نفس المجموعة الاجتماعية الاقتصادية. وبالتعامل مع زوج الآباء وزوج الأبناء، فإننا بالتالي ننظر إلى الفرق في كل زوج. وأن تأثير المتغيرات الأخرى كالوضع الاجتماعي والاقتصادي قد يزول Mitigated. في الواقع، نحن نقول إن كل المتغيرات الأخرى التي يمكن أن تحدد مشاهدة الإذاعة المرئية هي متساوية في كل زوج من الأزواج المعطاة، ولذلك، فإن العلاقة العائلية فقط تختلف بين هذه المتغيرات بحيث تسمح لنا بعزل تأثير هذه المتغيرات على المتغير التابع[10].

اختبار t لعينات تابعة لفرق المتوسط:

لتوضيح استخدام اختبار t لعينات تابعة (الأزواج) فإننا نورد المثال التالي:

مثال:

أُجْرِيَ مسحُ اجتماعيُّ لعدد 10 عائلات، أب من كل أسرة، وطفل من كل أسرة، وطُلب إليهم أن يقوم كل منهم بتسجيل كمية المشاهدة للإذاعة المرئية خلال فترة معينة من الوقت في مدونة يومية. وقد ظهرت البيانات في الجدول التالي:

جدول (13ـ2): مشاهدة الإذاعة المرئية وفقا لزوج الأسر

الأسرة	الدقائق المشاهدة من قِبل الأب	الدقائق المشاهدة من قِبل الطفل
1	23	45
2	25	56
3	43	73
4	26	53
5	21	27
6	29	34
7	32	76
8	23	21
9	25	54
10	21	43
	$\overline{X} = \frac{\Sigma x}{N} = 26.8$	$\overline{X} = \frac{\Sigma x}{N} = 48.2$

المصدر: George Argyrous, Statistics for Social & Health Research, op.cit. , P. 435

ولإجراء اختبار t لعينات تابعة يمكننا قلب الإجراءات التي تم اتباعها في اختبار t لعينتين مستقلتين:

1- حساب المتوسط الحسابي لكل عينة.

2- وبعد ذلك حساب الفرق بين وسطي العينتين.

وعند قلب هاتين الخطوتين لإجراء اختبار t لعينتين تابعتين يصبح الإجراء كالتالي:

1- حساب الفرق لكل زوج من الحالات (D)

2- وبعد ذلك حساب متوسط الفروق ($\overline{X}_D$)

ولكي نضع هذا الأمر بشكل أكثر دقة، فإنه يمكننا القول بأن اختبار t للعينات المستقلة تعبر عن الفرق بين المتوسطات، بينما يعبر اختبار t للعينات غير المستقلة عن متوسط الفروق.

والجدول رقم (13 – 3) يبين لنا الخطوة الأولى المتعلقة بإجراء اختبار للعينات غير المستقلة، وذلك من خلال حساب الفروق في كمية مشاهدة الإذاعة المرئية لكل زوج.

جدول (13 ـ 3)

الأسرة	الفرق في دقائق المشاهدة (D)
1	22 = 23 – 45
2	31 = 25 – 56
3	30 = 43 – 73
4	27 = 26 – 53
5	6 = 21 – 27
6	5 = 29 – 34
7	44 = 32 – 76
8	–2 = 23 – 21
9	29 = 25 – 54
10	22 = 21 – 43
متوسط الفرق $\overline{X}D = \frac{\Sigma D}{N} = 21.4$	

يمكننا ملاحظة أن متوسط الفرق مساوٍ للفرق بين المتوسطات، (وهو دائماً كذلك). أن السؤال الذي يمكن طرحه في هذا السياق هو، لماذا نقوم بهذا الإجراء البديل لحساب الفرق بين متوسطين؟. ومع أن متوسط الفرق سوف يكون دائماً مساوياً للفرق بين الوسطين، إلا أن التباين لن يكون الشيء نفسه، فالتباين حول متوسط الفرق يكون أصغر بكثير إذا قورن بالتباين حول الفرق بين الوسطين. والسبب في ذلك أنه من الممكن أن نعجز في رفض الفرق إذا ما تمت معاملته كفرق بين وسطين، عندما نكون قادرين على رفضه عندما نتعامل معه كمتوسط فرق.

وبالنظر إلى الجدول أعلاه فإننا نجد أن هناك في المتوسط فرقاً لكل الأزواج التي تشكل العينات. دعنا نفترض أنه في المجتمع ككل لا يوجد فرق في كمية المشاهدة للإذاعة المرئية بين الآباء وأبنائهم. وبالتالي يصاغ الفرض الصفري بالطريقة التالية:

$$H_O : \mu_D = 0$$

وعند إجراء عملية المعاينة من هذا المجتمع، فإننا قد نجد أحياناً أن أباً يشاهد الإذاعة المرئية أكثر مما يشاهدها ابنه، ولكن في أحيان أخرى قد نجد طفلاً يشاهد الإذاعة المرئية أكثر قليلاً عن عما يشاهده أبوه. ولكن إذا ما كان الفرض الصفري صحيحاً فإنه لا يوجد فرق في المتوسط فالفروق الموجبة سوف تلغي الفروق السالبة. بمعنى آخر، أنه من غير المعقول أن نتوقع أن التباين العشوائي قد يحدث أحياناً نتيجة لزيادة قليلة في الأسر التي يشاهد فيها الأب الإذاعة المرئية أقل من الطفل المناظر. أو العكس بالعكس، لذلك فإن فرق المتوسط بين العينات لن يكون صفراً (0)، فكلما كبر الفرق بين نتيجة العينة والنتيجة المتوقعة وهي صفر فرق المتوسط، مع ذلك، فإن احتمالاً ضئيلاً أن هذا سيكون راجعاً لتباين عشوائي، ولكن الأكثر احتمالاً أن هذا يعكس مبطن الفرق بين الآباء وأبنائهم.

في هذا المثال، فإن متوسط الفرق هو 21.4 دقيقة. فالسؤال المطروح في هذا الشأن هو هل هذا الفرق بين العينات يقودنا إلى رفض الفرض الذي مفاده لا يوجد فرق بين هذين المجتمعين؟

إن المعادلة المتعلقة بحساب اختبار t لفرق المتوسطات تكون:

$$t = \frac{\overline{X}_D}{S_D / \sqrt{N}}$$

حيث إن:

$$S_D = \sqrt{\frac{\Sigma D^2 - \frac{(\Sigma D)^2}{N}}{N-1}}$$

لاحظ أن (N) تشير إلى عدد الأزواج Number of Pairs وليس للعدد الكلي للحالات. في هذا المثال N = 10، بالرغم من أنه لدينا العدد الكلي الذي يصل إلى عشرين حالة، عشر حالات من الآباء، وعشر حالات من الأطفال.

إن درجة العينة لهذا المثال ستكون:

$$S_D = \sqrt{\frac{\Sigma D^2 - \frac{(\Sigma D)^2}{N}}{N-1}} = \sqrt{\frac{6410 - \frac{45.796}{10}}{10-1}}$$

$$= 14.2$$

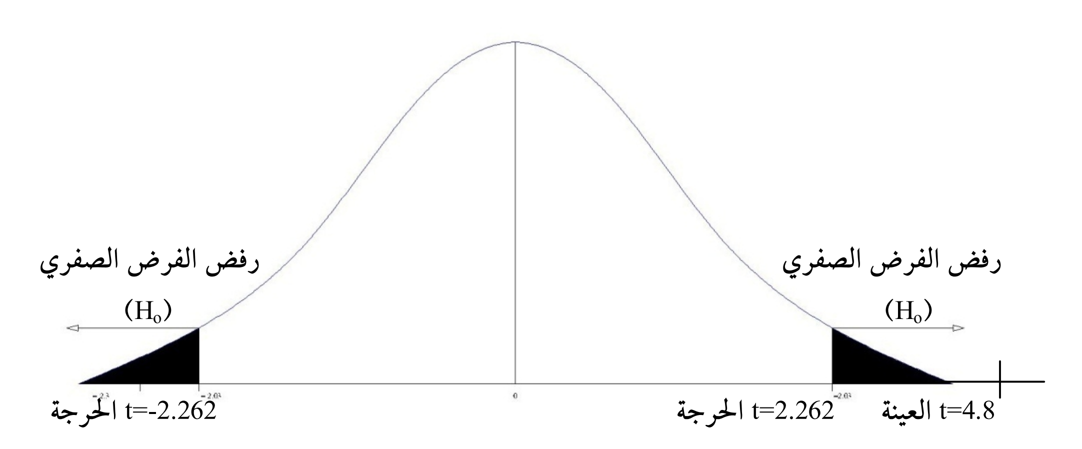
$$\text{العينة } t = \frac{\Sigma \overline{X}_D}{S_D / \sqrt{N}} = \frac{21.4}{14.2 / \sqrt{10}}$$

$$= 4.8$$

في هذا المثال لدينا 9 درجات حرية (10 أزواج - 1) على مستوى دلالة (α) 0.05، فالقيمة الحرجة لـِ t تساوي:

$$\text{الحرجة } t = \pm 2.262 \ (\alpha = 0.05, df = 9)$$

وإذا وضعنا هذه القيم في شكل رسم بياني (انظر الشكل رقم 13 - 6) فإننا نؤكد بأن الفرض الصفري قد تم رفضه. ومن هنا يمكننا القول، بأن هناك فرقاً دالاً في كمية مشاهدة الإذاعة المرئية بين الآباء والأبناء [11].



شكل رقم (13 ـ 6) الدرجة الحرجة ودرجة العينة

اختبار t لعينتين غير مستقلتين Paired-Samples t- test باستخدام SPSS :

الإجراء:

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Paired-Samples t- test ⇦ Copare means ⇦ Analyze

2- انقر على Minutes of TV Watched وبعد ذلك انقر على Parent

3- انقر على ◄ للصق المتغيرات في القائمة المحددة لِـ Paired Variables

4- انقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

T. Test

Paired Samples Statistics

Pair 1 Minutes of TV Watched	Mean	N	Std. deviation	Std. Error Mean
Parent Minutes of TV Watched	26.80	10	6.85	2.10
Child	48.20	10	18.05	5.71

Paired Samples Correlation

Pair 1 Minutes of TV Watched	N	Correlation	Sig.
Parent Minutes of TV Watched Child	10	.899	.024

Paired Samples test

Pair 1 Minutes of TV Watched	Paired differences					t	df	Sig (2-tailed)
	Mean	Std. Deviation	Std Error Mean	confidence %95 Interval of the difference				
				Lower	Upper			
Parent Minutes of TV Watched Child	−21.40	14.22	4.50	−31.57	−11.23	4.758	9	.001

المصدر: George Argyrous, Statistics for Social & Health Reoearch, op.cit. , P. 439

شكل (13 ـ 7) مخرجات SPSS لاختبار t لعينتين ثنائيتين

تفسير مخرجات لاختبار t لعينتين ثنائيين:

من خلال هذه المخرجات نلاحظ أن هناك فرقاً ذا دلالة في كمية مشاهدة الإذاعة المرئية بين الآباء والأبناء والأطفال، حيث بيَّن مربع Paired Samples Statistics أن درجات الوسط الحسابي عمود Mean لكل مجموعة (الآباء والأبناء) فقد وصلت إلى 26.80 الوسط الحسابي للآباء، و 48.20، الوسط الحسابي للأبناء؛ وبالتالي يمكننا القول بأن هناك زيادةً ذات دلالة في كمية مشاهدة الإذاعة المرئية بين الآباء والأبناء. وبالتالي تحصلنا على قيمة t مساوية لـِ4.758 بدرجة حرية 9، وعليه نرفض الفرض الصفري الذي مفاده لا يوجد فرق.

كذلك تمدنا هذه المخرجات مربع Paired Samples test بـِ 95% فترة ثقة لتقدير الفرق. فالحد الأعلى هو 11.23-، بينما الحد الأدنى وصل إلى 31.57-. ويمكن للباحث استخدام هذه المعلومات لإجراء الاختبار. ولما كانت الفترة لا تحتوي قيمة صفر (0)، فإنه يمكننا القول، بأن الفرق في المجتمع فيما يتعلق بكمية مشاهدة الإذاعة المرئية للآباء وأطفالهم ليس صفراً.

ولمعرفة حجم التأثير لاختبار t لعينتين ثنائيتين، يمكننا حساب eta^2 من خلال إحدى المعادلتين:

$$eta^2 = \frac{t^2}{t^2 + N - 1} \qquad (1)$$

أو

$$r^2 = \frac{t^2}{t^2 + df} \qquad (2)$$

من خلال المعادلة رقم (1):

$$= \frac{4.758}{4.758^2 + 10 - 1}$$

$$= \frac{22.638}{22.638 + 9}$$

$$= \frac{22.638}{31.638}$$

$$= 0.715$$

من خلال هذه القيمة المتوصل إليها وهي ($eta^2 = .715$) يمكننا القول، بأن هناك تأثيراً كبيراً مع وجود فرق جوهري في درجات اختبار كمية مشاهدة الإذاعة المرئية للآباء والأطفال.

مثال إضافي لزيادة التوضيح[11]:

يرغب مدرس في معرفة تأثير وسيلة جديدة في التعليم، على قدرة التلاميذ في إكمال مادة الرياضيات الأساسية. فقد اختار هذا المدرس خمسة من الطلاب، وطلب منهم إكمال اختبار مادة الرياضيات الأساسية. وبعد ذلك أدخل الوسيلة الجديدة في التعليم. وبعد مضي شهرُ كاملُ عاد المدرس فاختار نفس الطلاب الخمسة، وطلب منهم إكمال نفس الاختيار السابق؛ وقد جاءت النتيجة كالتالي:

جدول (13 ـ 3) نتيجة اختبار مادة الرياضيات الأساسية

التلاميذ	وقت إنهاء الاختبار (قبلي)	وقت إنهاء الاختبار (بعدي)
أحمد	7.3	6.8
إبراهيم	8.5	7.9
جمعة	6.4	6.0
عادل	9.0	8.4
هاني	6.9	6.5
المتوسط (Mean)	$\overline{X} = 7.62$	$\overline{X} = 7.12$

لقد تعامل هذا المدرس أساساً مع هذين الاختبارين كعينتين مستقلتين. إن متوسط الوقت للاختبار القبلي هو 7.62 دقيقة. في حين أن متوسط الوقت للاختبار البعدي قد وصل إلى 7.12 دقيقة. وباستخدامنا لاختبار t للعينة يقدر بـِ 0.75، وهي درجة غير دالة على مستوى 0.05. وبالرغم من هذه النتيجة فقد رأى هذا المدرس أن هذه النتيجة تبدو نتيجة واعدة. فاختبار الاستدلال الذي تم إجراؤه غير قادر على رفض احتمالية أن هذا

التحسن قد جاء نتيجة للصدفة. وعليه فقد قرر هذا المدرس أن يتخلى عن هذه الطريقة الجديدة في التدريس.

ولحسن الحظ، أن أحد زملاء هذا المدرس يعرف أكثر قليلاً حول الإحصاء، وقد أدرك ذلك بما أن نفس الطلاب يشكلون كل عينة فإن إجراء اختبار عينات غير مستقلة أصبح أمراً ضرورياً لتصميم البحث. وبالتالي قد تم التعامل من خلال البيانات بالنتائج التالية:

الخطوة الأولى: صياغة الفرض الصفري والفرض البديل:

$$H_0 : \mu_D = 0$$

$$H_i : \mu_D \neq 0$$

الخطوة الثانية: اختيار اختبار الدلالة:

في هذا المثال، نحن نقارن بين عينتين غير مستقلتين فيما يتعلق بفروق المتوسط، وعليه، فإننا نستخدم t لعينتين غير مستقلتين لفروق المتوسط.

الخطوة الثالثة: حساب درجة العينة:

ولحساب فرق المتوسط بين العينتين وربط ذلك بدرجة t يمكننا توليد الجدول التالي:

التلاميذ	وقت إنهاء الاختبار (قبلي)	وقت إنهاء الاختبار (بعدي)	الفرق D	D^2
أحمد	7.3	6.8	0.5	0.25
إبراهيم	8.5	7.9	0.6	0.36
جمعة	6.4	6.0	0.4	0.16
عادل	9.0	8.4	0.6	0.36
هاني	6.9	6.5	0.4	0.16
المجموع Sum			$\sum D = 2.5$	$\sum D^2 = 1.29$
المتوسط Mean			$\overline{X}D = 0.5$	

وبتعويض هذه البيانات لمعادلة الخطأ المعياري Standard error وبعد ذلك لمعادلة t.
نحصل على النتائج التالية:

$$S_D = \sqrt{\frac{\Sigma D^2 - \frac{(\Sigma D)^2}{N}}{N-1}} = \sqrt{1.29 - \frac{(2.5)^2}{5-1}}$$

$$= 0.1$$

$$t = \frac{\overline{X}D}{S_D / \sqrt{N}} = \frac{0.5}{0.1/\sqrt{5}}$$

$$= 11.8$$

الخطوة الرابعة: إيجاد الدرجة الحرجة والمنطقة الحرجة على مستوى دلالة 0.05 بدرجات حرية 4، وأن القيمة الحرجة لـِ t تساوي (2.776).

الخطوة الخامسة: اتخاذ القرار:

عندما تم حساب درجة t على أساس أن العينتين غير مستقلتين بدلاً من أنهما مستقلتين، فالنتيجة المتحصل عليها، نتيجة ذات دلالة بشكل واضح على مستوى 0.01. ومن هنا بإمكان هذا المدرس أن يرفض الفرضية التي مفادها أن التحسن جاء فقط من خلال الفرضية العشوائية.

أسئلة للمراجعة:

1- ما هي الافتراضات التي ينبغي أن تكون حول توزيعات المجتمعات قبل إجراء اختبار t للعينات المستقلة؟

2- النتائج التالية: اختبر فرق الدلالة مستخدماً اختباراً ثنائيَّ الجانبين بألفا α تساوي 0.05 (مفترضين تباينات متساوية Assuming Equal Population Variance

حجم العينة	الانحراف المعياري	المتوسط الحسابي		
35	14.2	72	عينة 1	a)
50	11	76.1	عينة 2	
100	0.9	2.4	عينة 1	b)
100	0.9	2.8	عينة 2	
120	80	450	عينة 1	C)
100	77	475	عينة 2	

3- يرغب باحث في معرفة تأثير مكان الإقامة على العمر الذي يبدأ فيه الناس بالتريض. فقد قسم الباحث العينة العشوائية للمبحوثين إلى فئتين: فئة الريفيين (91) فرداً، وفئة الحضريين (107) فرداً. وقد وجد أن السكان الريفيين قد بدأوا بالتدخين عند متوسط عمري يقدر بـِ 15.75 سنة، بانحراف معياري 2.3 سنة. في حين أن السكان الحضريين قد بدأوا التدخين عند متوسط عمري 14.63 سنة، بانحراف معياري 4.1 سنة. السؤال الذي يمكن طرحه هو: هل هناك فرق ذو دلالة بين هاتين الفئتين؟ (استخدم تقدير التباين المجمع).

4- المؤسسة العامة للمياه والصرف الصحي، ترغب في تقييم فعالية الإعلان الإرشادي حول ترشيد استهلاك المياه في مدينة بنغازي لغرض التقليل من استهلاك المياه غير المرشد. قبل الإعلان اختارت المؤسسة عينة عشوائية من (100) أسرة من أنحاء مختلفة من المدينة، وتم تسجيل استخدامات الأسر من المياه في الفترة الصباحية بمتوسط يصل إلى 87 لتراً، بانحراف معياري 15 لتر. وقد تم سحب (100) أسرة

عشوائياً بعد الإعلان، حيث وصل معدل الاستخدام للمياه 74 لتراً. هل يوجد فرق ذو دلالة؟ ما هي النتيجة التي يمكننا التوصل إليها حول الإعلان الإرشادي؟ هل يمكننا إجراء اختبار أحادي الجانب أم ثنائي الجانب؟ ما هي العوامل التي يجب أن نأخذها في الاعتبار عند اختيار الاختبار المناسب؟

5- أ) ما هو فرق المتوسط للأزواج العشرة المشاهدة:

زوج	المشاهدة (1)	المشاهدة (2)
1	12	15
2	10	13
3	8	13
4	14	14
5	12	18
6	15	13
7	14	18
8	9	9
9	18	11
10	13	14

ب) ما هو الخطأ المعياري (S_D).

ج) إجراء اختبار t لعينتين غير مستقلتين بمستوى دلالة $\alpha = 0.05$.

6- اختبر الفرضيات التالية مستخدماً البيانات التالية:

H_O	H_i	فرق المتوسط	SD	N	α
a) $\mu D = 0$	$\mu D \neq 0$	2.3	1.4	20	0.10
b) $\mu D = 0$	$\mu D < 0$	-3.2	20	41	0.05

7- شركة تريد أن تستقصي ما إذا كان التغير في منظومة العمل يمكن أن يكون دالاً في التحسين في مستويات الإنتاج. فقد اختارت هذه الشركة عشرة أماكن للعمل وتم قياس مستويات الإنتاج فيما يتعلق بعدد الوحدات المنتجة في الساعة. وبعد ذلك أدخلت هذه الشركة برنامجاً لهذه الأماكن الإنتاجية مانحة المنتجين حرية التصرف في شروط وبناء الوظيفة، وتم قياس مستويات الإنتاج بعد 6 أشهر. وقد جاءت النتائج كالتالي:

مكان العمل	الإنتاجية قبل التغيير	الإنتاجية بعد التغيير
1	120	165
2	121	154
3	145	120
4	112	155
5	145	164
6	130	132
7	134	154
8	126	162
9	137	130
10	128	142

- السؤال المطروح هو: هل إدخال البرنامج الجديد قد أدى بشكل دالٍ إلى تحسن مستويات الإنتاج (لاحظ صياغة الفرض البديل)؟

الهوامش والمصادر:

أولا الهوامش:

1- George Argyrous, Statistics for Social & Health Research, with a Guide to SPSS , SAGE Publications, London, 2001, P. 341.

2- Ibid. , P. 344.

3- Ibid. , P. 347.

4- Ibid. , PP. 349 - 350.

5- جولي بالانت (ت) خالد العامري، التحليل الإحصائي باستخدام برنامج SPSS، دار الفاروق للنشر والتوزيع، ط 2، مصر، 2009، ص 225.

6- Ibid. , P. 233.

7- Ibid. , P. 347.

8- Ibid. , PP. 351 - 352.

9- عبد الله عامر الهمالي، التقنيات الإحصائية ومناهج البحث، منشورات جامعة قاريونس، 2008 م.

10- George Argyrous , op.cit. , P. 434.

11- Ibid. , PP. 434 - 437.

12- Ibid. , PP. 440 - 442.

ثانيا: المصادر:

1- Frederick J. Gravetter and Larry B. Wallnau , Statistics for The Behavirol Sciences ,8th ed , Wadsworth Cengage Learning , USA, 2010.

2- George Argyrous, Statistics for Social & Health Research, with a Guide to SPSS , SAGE Publications, London, 2001.

3- جولي بالانت (ت) خالد العامري، التحليل الإحصائي باستخدام برنامج SPSS، دار الفاروق للنشر والتوزيع، ط 2، مصر، 2009.

4- عبد الله عامر الهمالي، التقنيات الإحصائية ومناهج البحث، منشورات جامعة قاريونس، 2008 م.

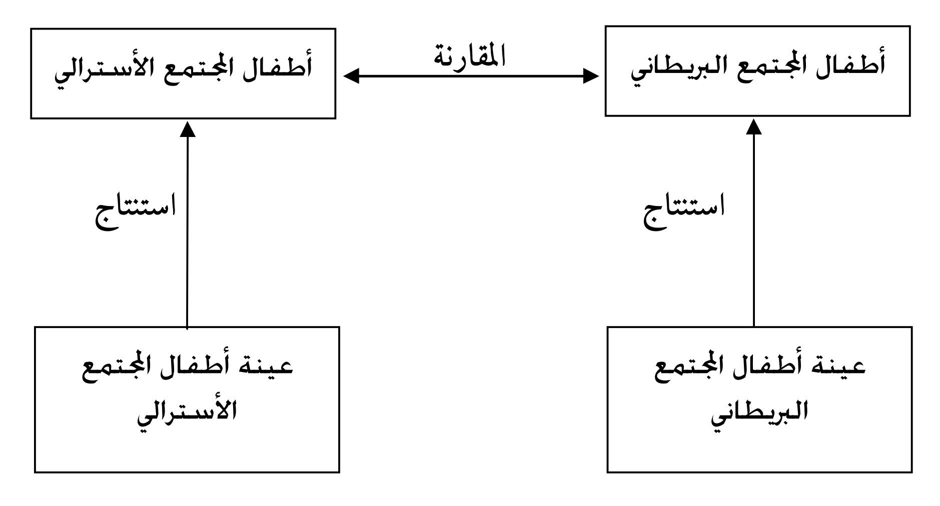
الفصل الرابع عشر
فروق الدلالة: اختبار F لأكثر من وسطين متساويين:
"تحليل التباين"

اختبار الفرض لأكثر من عينتين: الفكرة العامة:

تناولنا في الجزء الأول من الفصل السابق اختبار t لعينتين مستقلتين واختبرنا الافتراض بأن العينات قد تم سحبها من مجتمعات متوسطاتها متساوية:

$$Ho : \mu_1 = \mu_2$$

ولقد تعاملنا مع مثال حيث كانت العينة مكونة من عشرين طفلاً من المجتمع الأسترالي وعشرين طفلاً من المجتمع البريطاني. وقد تَّمَّ طرح سؤال على كل طفل في هاتين العينتين يطلب فيه من كل طفل أن يحدد عدد الدقائق التي يشاهد فيها الإذاعة المرئية في الليلة الواحدة. وقد تمت مقارنة العينتين من أجل اختبار الفرض الصفري Ho، الذي مفاده أنه لا يوجد فرق في متوسط كمية مشاهدة الإذاعة المرئية (التلفزيون) بين هؤلاء الأطفال في كلا المجتمعين (انظر الشكل التالي):

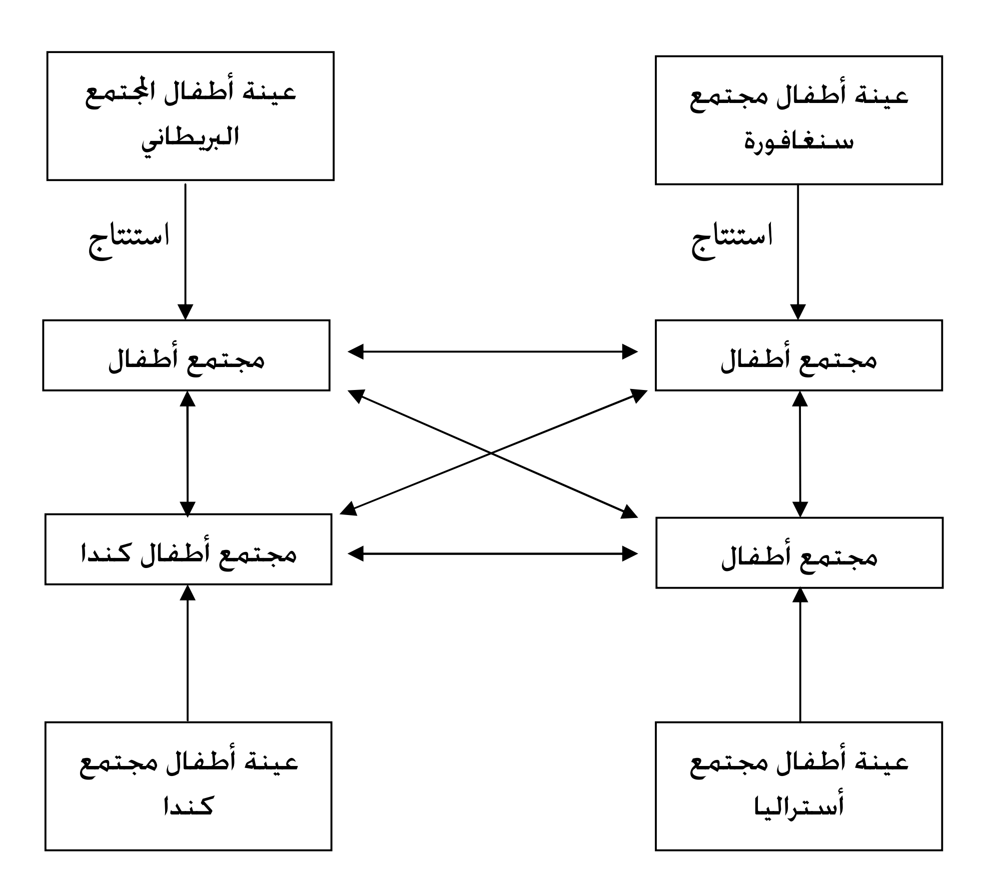


شكل (14 ـ 1) اختبار فرض لحالة عينتين

ويطلق على هذه العملية مسألة العينتين لأننا نتعامل مع عينتين من أجل الاستدلال حول كل مجتمع. إلا أننا في بعض الأحيان نتعامل مع مسألة أوسع بقليل. وبدلاً من مجرد المقارنة بين مجتمعين، فإننا قد نرغب في مقارنة كمية متوسط ما تمت مشاهدته في الإذاعة المرئية من قِبل الأطفال في عدة مجتمعات. على سبيل المثال، يمكن أن يكون لدينا عينة مؤلفة من عشرين طفلاً (20) تمثل المجتمعات التالية: أستراليا، بريطانيا، كندا وسنغافورة، ونود معرفة ما إذا كانت المتوسطات الحسابية لكل هذه المجتمعات متساوية:

$$Ho : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

ويطلق على هذه المسألة عينات K المستقلة K independent samples، حيث إن K تمثل أي عدد يزيد عن اثنين. وفي هذا المثال فإن K تساوي أربعة، ويمكننا توضيح هذا المثال بالشكل التالي:



المصدر: George Argyrous, Statistics for Social & Health Research,op. Cit , p. 356.

شكل (14 ـ 2) اختبار فرض لأكثر من عينتين

إن أحد الطرق للقيام بعملية المقارنة هي إجراء اختبار لكل عينتين متحدتين معاً مع العينات الأربعة بحيث يكون الحد الأعلى لهذه التركيبة هو 6، كما تم بيانه في الشكل أعلاه، فالأسهم تربط هذه المجتمعات بعضها ببعض:

المجتمع الأسترالي مع المجتمع السنغافوري

المجتمع الأسترالي مع المجتمع الكندي

المجتمع الأسترالي مع المجتمع البريطاني

المجتمع السنغافوري مع المجتمع الكندي

المجتمع السنغافوري مع المجتمع البريطاني

المجتمع الكندي مع المجتمع البريطاني

إذاً، يمكننا إجراء ستة اختبارات منفصلة لتقييم ما إذا كان هناك أي فروق دالة، تجدر الإشارة إلى أنه عندما نتعامل مع أكثر من عينتين، فإننا بذلك نستطيع اختبار متوسطات متساوية كلها مرة واحدة مستخدمين تحليل التباين لاختبار F (أنوفا ANOVA). والسبب وراء تفضيل استخدام تحليل التباين الأحادي على اختبارات t المتعددة يكمن في الوقوع في الخطأ من النوع الأول Type I error لسلسلة من اختبارات t ستكون أكبر من مستوى الدلالة المقررة لكل اختبار متعلق بـِ t. وعليه، إذا كان مستوى ألفا (α) لكل اختبار على حدة لـِ t هو 0.05، فإن فرصة الوقوع في الخطأ من النوع الأول لكل اختبارات t التي يمكن إجراؤها لعدد معين من العينات سوف يكون أكبر من 0.05 على الجانب الآخر، فإن اختبار أنوفا ANOVA، ينص على أن مستوى ألفا مساوٍ لخطر الوقوع في الخطأ من النوع الأول[1].

إن إجراء أنوفا لاختبارات الفرض الصفري الذي يشير إلى أن العينات سحبت من مجتمعات ذات متوسطات متساوية. فإذا كان الفرض الصفري صحيحاً، فإن العينات المسحوبة من مثل هذه المجتمعات سوف تكون قيمة متوسطاتها تقريباً متساوية. ففي مثال الأطفال ومعدل مشاهدة الإذاعة المرئية (التلفزيون)، فإن كل العينات لديها تقريباً متوسطات متساوية. وبطبيعة الحال، فإننا لا نتوقع أن تكون متوسطات العينة متساوية، حتى ولو كانت متوسطات المجتمع متساوية، حيث إن التباين العشوائي Random Variation سوف يؤثر على عملية المعاينة Sampling Process. إن السؤال الذي نطرحه هو ما إذا كانت الفروق بين العينات فروقاً متسقةً مع فرضية التساوي Assumption of Equality بين المجتمعات. إذا أخذنا بعين الاعتبار نتائج العينة الافتراضية للمجموعات الأربعة من الأطفال كما هو موضح في الجدول التالي:

جدول (14 ـ 1)

متوسط كمية مشاهدة الإذاعة المرئية في الليلة الواحدة (بالدقائق)

مشاهدة الإذاعة المرئية في الليلة الواحدة	المجتمع			
	كندا	أستراليا	بريطانيا	سنغافورة
المتوسط	127	166	187	203
الانحراف المعياري	27	29	30	26

المصدر: George Argyrous, Statistics for Social & Health Research,op. Cit , p. 357.

فإنه يمكننا ملاحظة أن هناك تبايناً واضحاً بين متوسطات هذه العينات الأربعة. وحقيقة الأمر، إذا قارنا القيم العليا بالقيم الدنيا لكل من كندا وسنغافورة، فإننا سنجد أن هناك فرقاً كبيراً في كمية متوسط مشاهدة الإذاعة المرئية. كما يمكن أيضاً ملاحظة (الصف المتعلق بالانحراف المعياري لكل عينة من العينات الأربعة) أن النتائج داخل عينة في كل مجتمع تتجمع معاً كما أشير إليها بانحرافات معيارية صغيرة نسبة إلى المتوسطات. بمعنى آخر، أنه توجد فروق واضحة من بلد إلى بلد آخر، ولكنها تتشابه داخل كل بلد. وفي ضوء هذه الإحصاءات الوصفية، فإنه بإمكاننا أن نبدأ بطرح تساؤل حول الفرض الذي مفاده أن المجتمعات التي من خلالها تم سحب هذه العينات هي مجتمعات لديها نفس المتوسط.

إن هذا المنطق يشبه تماماً ذلك المنطق المستخدم في تحليل التباين الذي يقوم على أساس مقارنة كمية التباين بين العينات مع كمية التباين داخل كل عينة من العينات. إذاً، بالرغم من أننا نرغب في معرفة الفرق بين المتوسطات، ففي واقع الأمر، فإن تحليل التباين يتعامل مع التباين الذي هو تربيع الانحراف المعياري.

تحليل التباين الأحادي: (اختبار F):

يمكننا الآن استخدام المفاهيم العامة لتحديد ما إذا كان هناك فرق دال بين الأطفال في مجتمعات مختلفة فيما يتعلق بكمية متوسط مشاهدتهم للإذاعة المرئية. ولحساب إحصاء الاختبار المناسب، فإننا نحتاج إلى بيان بعض المفاهيم الأساسية. أولها: كمية التباين الكلي

للدرجات المتعلقة بكل الحالات الثمانين (80) التي تم معاينتها. وقد تم قياس كمية التباين الكلي من خلال مفهوم يطلق عليه: مجموع التربيعات الكلية TSS. ويتم حساب التربيعات الكلية وفقاً للمعادلة التالية:

$$TSS = \sum (X - \overline{X})^2$$

إن قيمة مجموع التربيعات الكلية Tss يمكن أن تُقَسَّمَ إلى عنصرين، العنصر الأول هو كمية التباين داخل كل عينة، ويطلق عليه مجموع التربيعات داخل المجموعات (SSW). والعنصر الثاني هو كمية التباين بين كل عينة، ويطلق عليه مجموع التربيعات بين المجموعات (SSB):

$$\text{TSS} = \text{SSB} + \text{SSW}$$

إن كل عنصر من هذه العناصر المتعلقة بمجموع التربيعات الكلية TSS يمكن حسابها بالطريقة التالية:

$$SSW = \sum (X - \overline{X}_S)^2$$

$$SSB = \sum n_S (\overline{X}_S - \overline{X})^2$$

حيث إن: X_s متوسط العينة المعطاة.

: n_s عدد الحالات في العينة المعطاة.

تجدر الإشارة إلى أن هذه المعادلات تذكرنا بالمعادلة المتعلقة بالانحراف المعياري حيث إن هذه المعادلة تتضمن نفس الأساس بأن عملية التباين مرتبطة بكمية الفرق بين الدرجات الفردية والمتوسط، أما فيما يتعلق بمعادلة الانحراف المعياري، فإن هذه المعادلات المعَرَّفة Definitional Formulas، من الصعوبة بمكان العمل بها. وبشكل خاص، إذا أردنا أن نحسب المجموع الكلي للتربيعات TSS، فإنه من السهولة بمكان العمل بالمعادلة التالية:

$$TSS = \sum x^2 - n\bar{x}^2$$

ولكي أن نتحصل على مجموع التربيعات الكلية، فإننا بحاجة فقط لحساب إما SSW أو SSB وحينها نستخدم المعادلة التالية: TSS=SSB+SSW لحساب الأخرى. بمعنى آخر، إذا تم حساب قيمة TSS و SSB فإننا نستعيض بهذه الحسابات للمعادلة التالية لكي نتوصل إلى قيمة SSW:

$$SWW = TSS - SSW$$

ولكي نعرف كيف تم ذلك، دعنا نعيد العمل من خلال المثال السابق للعينات الأربعة المتعلقة بالعشرين طفلاً. ولإجراء هذه العمليات الحسابية بشكل جيد، يتطلب الأمر بناء جدول يحتوي على قائمة بيانات (انظر الجدول 2).

من خلال هذا الجدول سنجد أن درجة كل حالة من الحالات قد دونت، وأن العينات قد وضعت في أعمدة منفصلة. ومن خلال هذه المعلومات يمكننا حساب المتوسط الحسابي لكل عينة من هذه العينات، إضافة إلى المتوسط الحسابي لكل العينات موحدة.

كندا $\bar{X} = \frac{2546}{20} = 127.3$ دقيقة

أستراليا $\bar{X} = \frac{3317}{20} = 165.85$ دقيقة

بريطانيا $\bar{X} = \frac{3735}{20} = 186.75$ دقيقة

سنغافورة $\bar{X} = \frac{4063}{20} = 203.15$ دقيقة

جدول (14 ـ 2) لحساب تحليل التباين (ANOVA)

كندا		أستراليا		بريطانيا		سنغافورة	
X	X^2	X	X^2	X	X^2	X	X^2
89	7921	102	10,404	124	15,376	156	24,336
92	8464	120	14,400	135	18,225	165	27,225
95	9025	132	17,424	156	24,336	174	30,276
105	11,025	134	17,956	165	27,225	179	32,041
106	11,236	145	21,025	167	27,889	180	32,400
108	11,664	149	22,201	172	29,584	184	33,856
110	12,100	156	24,336	178	31,684	189	35,721
113	12,769	162	26,244	182	33,124	189	35,721
116	13,456	165	27,225	184	33,856	196	38,416
125	15,625	165	27,225	185	34,225	203	41,209
128	16,384	165	27,225	186	34,596	204	41,616
135	18,225	174	30,276	187	34,969	207	42,849
138	19,044	178	32,041	189	35,721	210	44,100
139	19,321	180	32,400	198	39,204	218	47,524
140	19,600	187	34,969	209	43,681	221	48,841
146	21,316	189	35,721	212	44,944	228	51,984
146	21,316	196	38,416	218	47,524	231	53,361
154	23,716	201	40,401	223	49,729	238	56,644
167	27,889	206	42,436	225	50,625	241	58,081
194	37,636	210	44,100	240	57,600	250	62,500
$\sum X = 2546$		$\sum X = 3317$		$\sum X = 3735$		$\sum X = 4063$	
$\sum X^2 = 337{,}732$		$\sum X^2 = 566{,}425$		$\sum X^2 = 714{,}117$		$\sum X^2 = 838{,}701$	

المصدر: George Argyrous, Statistics for Social & Health Research, with a Guide to SPSS , SAGE Publications, London , 2001 , p. 361.

$$\overline{X} = \frac{(2546+3317+3735+4063)}{80} = 170.8$$ دقيقة

وباستخدام هذه المعلومات يمكننا حساب TSS و SSB.

$$TSS = \sum X^2 - n\overline{x}^2$$

$$= (337,732 + 566,425 + 714,117 + 838,701) - (80)(170.8)^2$$

$$= 124,189$$

$$SSB = \sum n_S(\overline{x}_S - \overline{x})^2$$

$$= 20\,(127.3 - 170.8)^2 + 20\,(165.85 - 170.8)^2$$

$$+ 20\,(186.75 - 170.8)^2 + 20\,(203.15 - 170.8)^2$$

$$= 64,353$$

$$SSW = TSS - SSB = 124,189 - 64,353$$

$$= 59,836$$

إن إحصاء الاختبار الفعلي الذي يمكن استخدامه لتحديد ما إذا كان هناك فرق دال في تقدير التباين بين العينات وداخل العينات، نستخدم نسبة F (F.ratio). وتعني F.ratio نسبة تباين الاثنين: SSB و SSW وكل واحدة من هذه تصحح بدرجات حرية ملائمة:

العينة $F = \dfrac{SSB/K-1}{SSW/n-K}$

حيث إن: K تشير إلى عدد العينات.

وبتعويض الأرقام ذات الصلة لهذه المعادلة نحصل على:

العينة $F = \dfrac{SSB/K-1}{SSW/n-K}$

$$= \frac{64{,}353/4-1}{58{,}836/80-4}$$

$$= 27.25$$

وكما رأينا عند التعامل مع اختبارات Z، و T و X^2 لاحقاً، فإننا بحاجة إلى مقارنة الدرجة المحسوبة بالدرجة الاحتمالية من أجل الوصول إلى قرار يمكن من خلاله أن نقبل أو نرفض الفرض الصفري Ho.

ولإيجاد القيمة الحرجة يمكننا إحالة القارئ إلى جدول توزيع القيم الحرجة لِـ F على مستوى دلالة 0.05. انظر الجدول (14 – 4).

جدول (14 ـ 3): تحليل التباين الأحادي

(دقائق مشاهدة الإذاعة المرئية في الليلة الواحدة)

مصادر التباين	درجة الحرية	مجموع المربعات	متوسط مجموع المربعات	اختبار (F)	درجة الدلالة 0.05 (α)
بين المجموعات	3	64353.438	21451.146	27.246	.000
داخل المجموعات	76	59835.050	787.303		
المجموع الكلي	79	124188.488			

وللحصول على القيم الحرجة لاختبار F ينبغي على الباحث أن يراعي ثلاثة عوامل أساسية:

1- درجات الحرية لتقدير التباين بين العينات. وللحصول على درجة الحرية: عدد العينات ناقص واحد. ويظهر هذا في البسط المتعلق بِـ F- ratio:

$$Df_b = k - 1 = 4\text{-}1$$

$$= 3$$

$$Df_w = n - k = 80 - 4$$

$$= 76$$

جدول (14ـ4) القيم الحرجة لتوزيعات F (ألفا 0.05 = α)

درجات الحرية لتقدير التباين بين عينات K-1										
n-k	1	2	3	4	5	6	7	8	9	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.84	8.81	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.63
5	6.16	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.36
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	1.51
50	4.03	3.18	2.79	2.56	2.38	2.29	2.20	2.13	2.07	1.44
60	4.00	3.15	2.76	2.52	2.37	2.25	2.17	2.10	2.04	1.39
80	3.96	3.11	2.72	2.48	2.33	2.19	2.12	2.05	1.99	1.32
100	3.94	3.09	2.70	2.46	2.30	2.17	2.10	2.03	1.97	1.28
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.25
α	3.84	2.99	2.60	2.37	2.21	2.09	2.01	1.94	1.88	1.00

من خلال جدول توزيع القيم الحرجة لِـ F على مستوى دلالة 0.05 يمكننا أن نلاحظ أنه لا يوجد صف متعلق بدرجة الحرية داخل درجة تساوي 76. في الحقيقة أن مدى كل القيم قد قفزت بعد أول 30؛ ويرجع السبب في ذلك، إلى أن درجات الحرية لا تقل كثيراً عن زيادة مقدار في درجة الحرية بعد الثلاثين. وعندما لا تكون لدينا درجات حرية تظهر في الجدول، فإننا نشير إلى أقرب قيمة تظهر في الجدول من الرقم المرغوب فيه: في هذا المثال، إن أقرب قيمة دون 76 تظهر في الجدول هي عند الرقم 60. وعليه، فإن القيمة الحرجة لِـ F تكون في المثال هي 2.76. ومن هنا نرفض الفرض الصفري Ho الذي مفاده لا فرق باعتبار أن قيمة F المحسوبة أكبر من قيمة F المتوقعة:

نرفض الفرض الصفري عندما تكون (F المتوقعة < F العينة).

وعليه عند رفضنا للفرض الصفري فإنه على الأقل واحد من مجتمعات هؤلاء

الأطفال تختلف عن الأخريات. لاحظ التعبير الخاص للنتيجة: على "الأقل" واحد من هذه المجتمعات يختلف عن باقي المجتمعات الأخرى.

إن اختبار F في حد ذاته لا يخبرنا أي المجتمعات، وكم كمية الاختلاف. وبكل وضوح إذا كانت هناك فروق، إذاً على الأقل يجب أن تشمل المجموعات التي لديها عينة بمتوسطات عليا ودنيا في هذا المثال ستكون كندا وسنغافورة على التوالي.

وكما أوضحنا سابقاً برفضنا للفرض الصفري بعد إجراء اختبار F فإننا بذلك قد قررنا أن واحداً على الأقل من هذه المجتمعات لديه متوسط حسابي غير مساوٍ لمتوسطات المجتمعات الأخرى. وعليه فإن اختبار F في حد ذاته لا يخبرنا أي هذه المجتمعات يكون مختلفاً. ولكي نحدد أياً من المجتمعات يختلف عن الأخرى، فالباحث يمكنه أن يلجأ إلى استخدام Post - Hoc Comparisons المتوفرة كخيار في برنامج SPSS[2].

إجراء اختبار F باستخدام Spss:

1- من القائمة الموجودة في الجزء العلوي من الشاشة انقر فوق:

Analyze ⇦ **Compare Means** ⇦ **one Way ANOVA**

2- انقر فوق المتغير التابع **Minutes of TV Watched** في قائمة المتغير.

3- انقر على ◄ مشيراً إلى الصندوق أسفل **Dependent List**. هذه القائمة تقود إلى لصق **Minutes of TV Watched** في القائمة المحددة للمتغير التابع الذي يستخدم لمقارنة العينات.

4- انقر على **Country of Residence** في القائمة.

5- أنقر على ◄ مشيراً إلى الصندوق أسفل: **Factor** تقود هذه العملية إلى لصق **Country of Residence** في القائمة المحددة لمتغير **Factor** الذي يشكل العينات المستهدفة للمقارنة.

6- انقر على ok.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

ANOVA

ONE Way Minutes of TV Watched Per night

	Sum of Squares	dF	Mean Square	F	Sig.
Between groups	64353.438	3	21451.146	27.246	.000
Within groups	59835.050	76	787.303		
TOTAL	124188.488	79			

المصدر: George Argyrous, op. , p. 361.

شكل (14 ـ 3) مخرجات لتحليل التباين الأحادي

تفسير مخرجات SPSS لتحليل التباين:

بالنظر إلى مخرجات SPSS لتحليل التباين يمكننا أن نلاحظ في المربع أعلاه مجموع التربيعات بين المجموعات Sum of Squares Between Groups، ومجموع التربيعات داخل المجموعات Sum of Squares Within Groups، ومجموع التربيعات الكلية The Total Sum of Squares في العمود الأول في مربع ANOVA معاً مع درجة الحرية المناسبة (df) في العمود الثالث. ومن خلال هذه المخرجات نجد أن قيمة F تساوي 27.246 (العمود 5)، وأن درجة الدلالة عمود .Sig تساوي 000.، ولا تعني هذه الدلالة أن احتمالية الحصول على نسبة 27.25 لـِ F تساوي صفراً (0)، ذلك أن برنامج SPSSيقرب الاحتمالية إلى ثلاث درجات عشرية decimal Places3 ، وعليه، فإن هذه الاحتمالية تقرأ أقل من 5 في 10.000.

ولمّا كانت القيمة في هذا المثال (Sig. = 0.000)، وهي أقل من 05. مما يدل على وجود فروق دالة بين المجموعات.

حجم التأثير:

ولحساب حجم التأثير الذي يعطي مؤشراً لحجم الفروق بين المجموعات (وليس مجرد تحديد ما إذا كان الفارق قد حدث على سبيل الصدفة أم لا) يستخدم إحصاء Eta^2

إيتا تربيع. ولما كانت مخرجات SPSS لا تقوم بحساب Eta^2 في اختبار F، فالباحث يمكنه حسابها من المخرجات المرتبطة باختبار F، وفقاً للمعادلة التالية:

$$Eta^2 = \frac{\text{مجموع التربيعات بين المجموعات}}{\text{المجموع الكلي للمربعات}}$$

$$Eta^2 = \frac{SSb}{TSS} = \frac{64353.438}{124188.488}$$
$$= 0.518$$

من هنا نجد أن حجم التأثير يبلغ 0.518 أو (52 %) وهو تأثير كبير نسبياً حسب إرشادات كوهين، مع وجود فرق دال في متوسط عدد الدقائق المشاهدة في الليلة الواحدة.

مثال لزيادة التوضيح [3]:

في دراسة قام بها أحد الباحثين على ثلاثة أطفال لمقارنتهم في القدرة على القراءة. وقد طلب من كل طفل قراءة 12 واجباً مع ضبط الأخطاء خلال قراءة كل واجب من هذه الواجبات الاثنى عشر، ويود الباحث معرفة هل هؤلاء الأطفال يختلفون في قدرات القراءة؟ ويوضح الجدول التالي النتائج المتحصل عليها:

جدول (14-4) عدد الأخطاء لكل طفل

رقم الواجب	زيد	عمرو	خالد
1	8	15	12
2	6	9	6
3	14	20	8
4	9	15	9
5	14	6	10
6	8	9	14
7	12	17	16
8	19	12	5
9	6	6	18
10	11	3	21
11	8	13	15
12	15	5	11

من خلال هذه البيانات، هل يمكننا القول بأن هناك فروقاً ذات دلالة بين هؤلاء الأطفال الثلاثة في قدرات القراءة؟

للإجابة على هذا السؤال يمكننا إجراء اختبار F وذلك باتباع الخطوات التالية:

الخطوة الأولى: صياغة الفرض الصفري والفرض البديل:

Ho: متوسط عدد الأخطاء التي ارْتُكِبَتْ من قِبل كل طفل من هؤلاء الأطفال هي متوسطات متساوية: $Ho: \mu_1 = \mu_2 = \mu_3$

Hi: متوسط عدد الأخطاء التي ارْتُكِبَتْ من قِبل كل طفل من هؤلاء الأطفال ليست كلها متوسطات متساوية: $Hi: \mu_1 \neq \mu_2 \neq \mu_3$

الخطوة الثانية: اختيار اختبار الدلالة:

إن السؤال البحثي الذي يرغب الباحث فيه هو متوسط عدد الأخطاء التي تم قياسها على مستوى المقياس ذي المسافات والنسبي. وعليه فالرغبة تنحصر في مقارنة المتوسطات وبيان ما إذا كانت هذه المتوسطات متساوية. ولما كانت لدى الباحث ثلاث عينات لذلك ستكون المقارنة عبر أكثر من عينتين. وأن الاختبار الملائم لهذه المسألة البحثية هو تحليل التباين (اختبار F لمتوسطات متساوية).

الخطوة الثالثة: حساب درجة العينة:

لإجراء تحليل التباين ANOVA، فإنه من المفيد أن يوجد جدول يحتوي على الحسابات المرغوبة (انظر الجدول 5)، ومن هذه المعلومات نقوم بحساب المتوسط الحسابي لكل عينة من العينات الثلاثة، وبعد ذلك تجمع المتوسطات لكل هذه العينات.

زيد $\overline{X} = \frac{130}{12} = 10.8$

عمرو $\overline{X} = \frac{140}{12} = 11.7$

خالد $\overline{X} = \frac{145}{12} = 12.1$

جدول (14 ـ 5) حساب تحليل التباين ANOVA

زيد		عمرو		خالد	
X	X^2	X	X^2	X	X^2
8	64	15	225	12	144
6	36	9	81	6	36
14	196	20	400	8	64
9	81	15	225	9	81
14	196	6	36	10	100
8	64	9	81	14	196
12	144	17	289	16	256
19	361	12	144	5	25
6	36	6	36	18	324
11	121	13	169	21	441
8	64	13	169	15	225
15	225	5	25	11	121
$\Sigma X = 130$	$\Sigma X^2 = 1588$	$\Sigma X = 140$	$\Sigma X^2 = 1880$	$\Sigma X = 145$	$\Sigma X^2 = 2013$

$$\overline{X} = \frac{130 + 140 + 145}{36} = 11.5$$

وتمثل هذه العمليات الحسابية الإحصاء الوصفي لبيانات العينة. وبوضوح بأنه يوجد فرق بين العينات فيما يتعلق بمتوسط عدد الأخطاء المرتكبة. هل هذه الأخطاء المرتكبة راجعة إلى تباين عشوائي عند عملية المعاينة من مجتمعات لا يوجد فرق فيما بينها؟

ولكي نقرر ذلك، يمكننا بادئ ذي بدء حساب TSS و SSB:

$$TSS = \Sigma X^2 - n\overline{x}^2 = (1588 + 1880 + 2013) - 36(11.5)^2$$

$$= 720$$

$$SSB = \Sigma ns(\bar{x}_S - \bar{x})^2$$

$$= 12\ (10.8 - 11.5)^2 + 12\ (11.7 - 11.5)^2 + 12(12.1 - 11.5)^2$$

$$= 10.7$$

$$SSW = TSS - SSB = 720 - 10.7$$

$$= 709.3$$

ومن خلال هذه البيانات يمكننا حساب قيمة F الإحصائية وذلك باستخدام اختبار الدلالة:

$$F_{\text{العينة}} = \frac{SSB/K-1}{SSW/n-K}$$

$$= \frac{10.7/3-1}{709.3/36-3}$$

$$= 0.25$$

الخطوة الرابعة: إيجاد الدرجة الحرجة والمنطقة الحرجة:

من خلال جدول توزيع F الحرجة، فإن K=2 و N=36

وعليه، فإن القيمة الحرجة لـِ F:

$$F_{\text{الحرجة}} = 3.32\ (\alpha = 0.05)$$

الخطوة الخامسة: اتخاذ القرار:

إذا أردنا رسم درجة العينة ودرجة القيمة الحرجة في الشكل رقم (14 – 3)، فإننا بذلك لا يكون في مقدورنا رفض الفرض الصفري.

حجم التأثير: ولمعرفة مؤشر حجم الفروق بين هذه المجموعات (ليس مجرد تحديد ما إذا كان الفارق قد حدث بالصدفة أم لا)، يمكننا استخدام إحصاء Eta^2 إيتا تربيع التي يمكن حسابها وفقاً للمعادلة التالية:

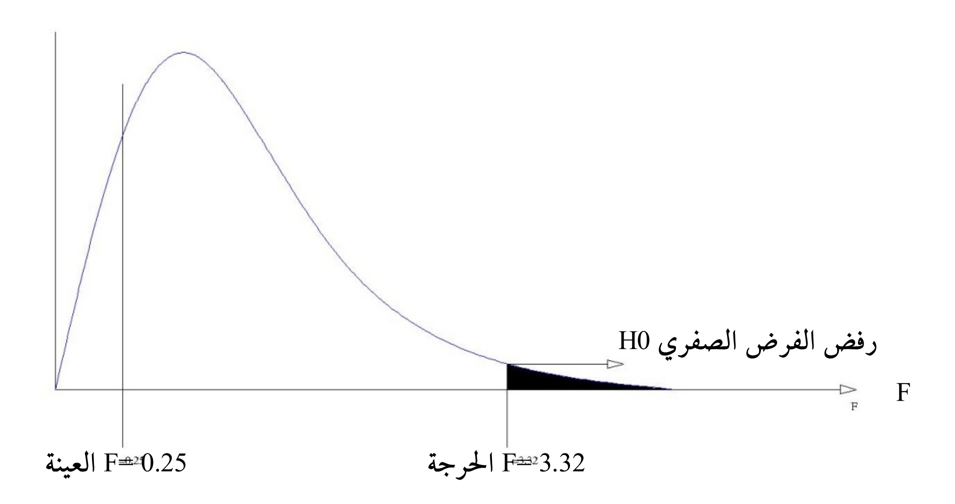
$$\text{إحصاء إيتاء تربيع} = \frac{\text{مجموع التربيعات بين المجموعات}}{\text{المجموع الكلي للمربعات}}$$

$$\text{Eta}^2 = \frac{ssb}{Tss}$$

$$= \frac{10.7}{720}$$

$$= 0.015$$

من هنا نجد أن حجم التأثير يبلغ 0.015 في المائة وهو تأثير ضئيل حيث يشير إلى أنه لا يوجد فرق جوهري في متوسط عدد الأخطاء لدى عينات هؤلاء الأطفال.



شكل (14-3) درجات العينة والحرجة

أسئلة للمراجعة:

1- لقد تمت المقارنة بين أربع جمعيات للرعاية في إطار متوسط عدد الحالات التي يقدمها موظفو هذه الجمعيات خلال شهر. ويهدف البحث للإجابة عن السؤال ما إذا كان الفرق بين هذه الجمعيات وفقاً لعبء العمل ذا دلالة.

أ- اشرح لماذا يستخدم تحليل التباين ANOVA لسبر غور هذه المسألة.

ب- بين الفرض الصفري لهذه المسألة لفظياً وجبرياً.

ج- من النتائج الافتراضية احسب قيمة F واتخذ القرار المناسب حول الفرض الصفري (Ho) بمستوى دلالة (α = 0.05).

التباين	مجموع التربيعات	درجة الحرية
بيــن الجمعيات	50	4
داخل الجمعيات	7210	110
المجموع الكلي	**7260**	**114**

2- مُحَاضِرٌ بالجامعة يستخدم طرق تدريس مختلفة في ثلاثة فصول دراسية منفصلة. وكان الهدف من وراء هذا الاستخدام تقييم الفعالية النسبية لهذه الطرق باختبار فرق الدلالة بين هذه الفصول. والدرجات النهائية تظهر في الجدول التالي:

الطريقة (أ)	الطريقة (ب)	الطريقة (ج)
21	28	19
19	28	17
21	23	20
24	27	23
25	31	20
20	38	17
27	34	20
19	32	21
23	29	22
25	28	21
26	30	23

المطلوب:

أ- احسب المتوسط الحسابي، والانحراف المعياري لكل عينة من العينات الثلاث. هل تستطيع أن تتوقع من هذه الإحصاءات الوصفية النتيجة المتحصل عليها من خلال إجراء أنوفا ANOVA على هذه البيانات.

ب- أَجْرِ تحليل ANOVA لتقييم هذه التوقعات.

3- البيانات التالية، بيانات افتراضية حول عينة مؤلفة من 20 طفلاً من مجتمع الولايات المتحدة الأمريكية تبين عدد الدقائق التي يشاهدها هؤلاء الأطفال في الليلة الواحدة:

195	184	165	162	168	196	217	190	212	232
204	205	217	210	230	197	180	192	190	198

المطلوب:

أ- كيف تُؤثر إضافة هذه العينة على اختبار تحليل التباين ANOVA على أطفال مجتمعات: أستراليا، بريطانيا، كندا، وسنغافورة، على عدد درجات الحرية؟

ب- ما النتيجة التي يمكنك التوصل إليها فيما يتعلق بكمية مشاهدة الأطفال للإذاعة المرئية من بلدان مختلفة؟

هوامش والمصادر:

أولا: الهوامش:

1- George Argyrous, Statistics for Social & Health Research, with a Guide to SPSS , SAGE Publications, London, 2001, P. 357.

2- Ibid. , PP. 364 - 365.

3- Ibid. , PP. 367 - 370.

ثانيا: المصادر:

1- George Argyrous, Statistics for Social & Health Research, with a Guide to SPSS , SAGE Publications, London, 2001.

2- جولي بالانت (ت) خالد العامري، التحليل الإحصائي باستخدام برنامج SPSS، ط 2، دار الفاروق للنشر والتوزيع، مصر، 2009.

الجزء الخامس

الإحصاءات الوصفية المتعددة

- الفصل الخامس عشر: التوسع في جداول التقاطع: إضافة متغيرات التحكم
- الفصل السادس عشر: الارتباط الجزئي والانحدار المتعدد

الفصل الخامس عشر

التوسع في جداول التقاطع: إضافة متغيرات التحكم

مقدمة:

تناولنا في الفصل السادس تحليل العلاقة بين متغيرين، وقد تم افتراض أن أي علاقة تمتْ مشاهدتها في البيانات بين متغيرين هي ناشئة عن علاقة بسيطة ومباشرة. إن التطابق القوي في الجداول الثنائية، على أية حال، لا يعني بالضرورة أن هذه العلاقة البسيطة المباشرة موجودة واقعياً فالأمر يتعلق فقط، بالكيفية التي قمنا بها في تفسير هذه البيانات. ويمكننا القول، إن هناك علاقات أكثر تعقيداً تخفيها البيانات، لكننا لم نبذل جهداً كافياً لاكتشاف هذه العلاقات.

إن أسهل طريقة للتوسع في العلاقة، هي اكتشافها في جدول التوافق وذلك من خلال النظر إلى احتمالية تأثير متغير ثالث على العلاقة الثنائية الأصلية. واعتماداً على مخرجات هذا التوسع، فإنه يمكننا أن نعدل من نموذج العلاقة بين المتغيرين الأصليين آخذين في الاعتبار تأثير المتغير الثالث.

هناك ثلاث نتائج يمكن الوصول إليها عند إدخال المتغير الثالث في عملية التحليل:

1- النتيجة الأولى أنه ليس للمتغير الثالث أي تأثير على العلاقة الأصلية المباشرة.

2- قد يكون لإدخال المتغير الثالث تأثير في العلاقة الموجودة.

3- أو أن هناك علاقة مشروطة[1].

في هذا الفصل سوف نقوم باستقصاء هذه النتائج المحتملة، وذلك من خلال إعطاء أمثلة لكل نمط من أنماط هذه العلاقة.

1- العلاقة المباشرة:

دعنا ننظر إلى النقطة الأولى التي مفادها أنه لا يوجد أي تأثير لمتغير التحكم في العلاقة الثنائية الأصلية. ويبرهن لنا، هذا، على أن النموذج المباشر البسيط هو النموذج الملائم لوصف العلاقة.

على سبيل المثال، يمكننا إقامة علاقة بين الدخل ومشاهدة الإذاعة المرئية كما هي مرتبة في جدول التقاطع.. (انظر جدول 15 – 1).

جدول (15 - 1): مشاهدة الإذاعة المرئية حسب مستوى الدخل

مشاهدة الإذاعة المرئية	الدخل		المجموع
	منخفض	عال	
منخفض	115 % 57	95 % 32	210
عال	88 % 43	204 % 68	292
المجموع	203	299	502
Gamma = 0.47 = جاما			

إن العلاقة الأصلية في هذا النموذج بين هذين المتغيرين هي علاقة مباشرة كما في الشكل التالي:

مستوى الدخل ← مشاهدة الإذاعة المرئية

شكل (15 ـ 1) علاقة مباشرة

إن النموذج النظري الذي اعتمدنا عليه يوضح أن الدخل يؤثر مباشرة في كمية مشاهدة الإذاعة المرئية للشخص، وأن الإحصاءات تؤكد لنا أن هذه العلاقة الموجبة متوسطة إلى قوية، وعندما نجادل أن هناك علاقة مباشرة بين هذين المتغيرين فإننا بشكل فعّال نجادل في أن العلاقة سوف تكون نفسها بغض النظر عن أي متغير آخر يمكن أن يسبب تباين الحالات فيما بينها، في هذا المثال نعتقد أن الدخل يؤثر في مشاهدة الإذاعة المرئية بنفس الطريقة وبنفس الدرجة بغض النظر عن أي متغير آخر يمكنه أن يسبب التباين في الحالات مثل النوع، العمر... الخ.

إن هذه العلاقة الثنائية المباشرة على - أي حال - يمكن أن تظهر لنا أنها علاقة بسيطة. بالتأكيد توجد متغيرات أخرى قد تؤثر على كمية مشاهدة الإذاعة المرئية. فقد يشعر باحث آخر على سبيل المثال، بأن (المستوى التعليمي) يؤثر أيضاً في مشاهدة الأفراد للإذاعة المرئية. ولتقييم احتمالية تأثير هذا المتغير الجديد (المستوى التعليمي) أن له تأثير على العلاقة المشاهدة بين الدخل وكمية مشاهدة الإذاعة المرئية، يمكننا تقسيم العينة إلى مجموعتين فرعيتين:

المجموعة الأولى: تمثل أولئك الذين لم يكملوا التعليم الثانوي، والمجموعة الثانية: تمثل أولئك الذين أكملوا بعضاً من التعليم العالي. ومن الناحية الفنية فإننا قد أدخلنا متغير المستوى التعليمي كمتغير للتحكم. إن متغير التحكم هو ذلك المتغير الذي يحلل البيانات إلى مجموعات فرعية تستند على فئات متغير التحكم، إن تأثير متغير التحكم يؤدي بنا إلى توليد جداول متقاطعة منفصلة لكل المجموعات الفرعية والتي تم تحديدها من خلال متغير التحكم، في هذا المثال: دعنا نأخذ أولاً تلك (الحالات التي ليس لها تعليم عالٍ) ونصمم جدول تقاطع بين دخولهم ومشاهدتهم للإذاعة المرئية متغاضين عن تلك الحالات التي أكملت بعضاً من التعليم العالي، وبعد ذلك نأخذ فقط الحالات التي لديها بعضاً من

التعليم العالي ونصمم جدول تقاطع بين دخولهم ومشاهدة الإذاعة المرئية متغاضين عن أولئك الذين ليس لديهم تعليم عال.

إن ما قمنا به من توليد لهذه الجداول يطلق عليه الجداول الجزئية[2]. ويمكننا أن نولد من الجداول الجزئية العدد الذي نريده كفئات لمتغير التحكم (انظر جدول 15-2، 15-3) فإننا في هذين الجدولين قد أدخلنا المستوى التعليمي كمتغير للتحكم وتم قياس التعليم بفئتين فقط. وعليه يمكننا أن نولد جدولين جزئيين. (إذا كان لدينا ثلاث فئات لمتغير التحكم مثل: لا تعليم جامعي، بعض التعليم الجامعي، كثير من التعليم الجامعي في هذه الحالة يمكن توليد ثلاثة جداول جزئية).

جدول (15 ـ 2)

مشاهدة الإذاعة المرئية حسب مستوى الدخل

التحكم في مستوى التعليم (الذين ليس لديهم تعليم عالٍ)

مشاهدة الإذاعة المرئية	الدخل		المجموع
	منخفض	عال	
منخفض	78 A 57 %	22 B 31 %	100
عال	58 C 43 %	48 D 69 %	106
المجموع	136	70	206
جاما = 0.49 = Gamma			

جدول (15 ـ 3)

مشاهدة الإذاعة المرئية

التحكم في مستوى التعليم (الذين أكملوا بعضاً من التعليم)

مشاهدة الإذاعة المرئية	الدخل		المجموع
	منخفض	عال	
منخفض	37 % 55	73 % 32	110
عـــــال	30 % 45	156 % 68	186
المجموع	67	229	296
جاما = 0.45 = Gamma			

وبهذه النتيجة يمكننا القول إن العلاقة الأصلية قد ولَّدت لنا تقريباً علاقة متساوية في كل جدول جزئي، وإن قيمة جاما لكل واحد من هذين الجدولين الجزئيين تقرب كثيراً من قيمة جاما في الجدول الأصلي قبل التحكم في مستوى التعليم. بمعنى آخر، بغض النظر عن المستوى التعليمي فإن العلاقة بين الدخل ومشاهدة الإذاعة المرئية ما زالت قائمة، إن العلاقة المباشرة التي شاهدناها في البداية لازالت باقية على ما هي عليه حتى بعد التحكم في المتغير الثالث. وبغض النظر عن كيف تتباين الحالات طبقاً لمستوى الدخل، فالعلاقة الثنائية المباشرة تبقى الشيء نفسه، وعليه فلن نغير من نموذجنا الأصلي الذي يوضح أن هناك علاقة مباشرة بين الدخل ومشاهدة الإذاعة المرئية[3].

إجراءات توليد جداول التقاطع مع التحكم في المتغيرات باستخدام SPSS:

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Analyze/ descriptive Statistics / Crosstabs (يعطيك مربع جداول التقاطع).

2- انقر على TV Watching (مشاهدة الإذاعة المرئية).

3- انقر على ◄التي تشير إلى القائمة المحددة المعنونة: Row (s) (الصف أو الصفوف).

نقوم بلصق (Pasts) TV Watching في القائمة المحددة للصف أو الصفوف Row (s)

4- انقر على Income.

5- انقر على ◄ التي تشير إلى القائمة المحددة المعنونة Column(s)

نقوم بلصق (Pasts) Income في القائمة المحددة للأعمدة Column(s).

6- انقر على مستوى التعليم Education Level.

7- انقر على ◄ التي تشير إلى القائمة المحددة أسفل Layer 1 of 1، نقوم بلصق (Pasts) Education Level في القائمة المحددة التي تحتوي المتغير المتحكم The Control Variable جدول التقاطع سوف يولد كل قيمة للمتغير في القائمة.

8- انقر على زر Statistics وقم باختيار Gamma هذه العملية ستولد قيمة جاما Gamma لكل جدول جزئي.

9- انقر على الخلايا Cells واختر نسب العمود Column Percentages.

10- انقر على Ok.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

Cross tabs TV Watching "Income" Education Level Crosstabulation

Education Level			Income		ToTAL
			Low	High	
No Post. Secondary	TV Watching Low	Count	78	22	100
		% Within Income	57.4 %	34.4 %	48.5 %
	High	Count	58	48	106
		% Within Income	42.6 %	88.6 %	51.5 %
	TOTAL	Count	136	70	206
		% Within Income	100 %	100 %	100 %
Post. Secondary	TV Watching Low	Count	37	73	110
		% Within Income	55.2 %	31.9 %	37.2 %
	High	Count	30	156	186
		% Within Income	44.8 %	68.1 %	62.8 %
	TOTAL	Count	67	229	296
		% Within Income	100 %	100 %	100 %

Symmetric Measures

Education Level			Value	Asymp. std. Error[a]	Approx.T[b]	Approx. Sig.
No Post. Secondary	ordinal by Ordinal	GAMMA	.492	.118	3.657	.000
	N of Valid cases		206			
Post. Secondary	ordinal by Ordinal	GAMMA	.450	.113	3.317	.001
	N of Valid cases		296			

a. Not assuming The null hypothesis.

b. Using The asymptotic standard Error assuming the null hypothesis.

المصدر: George Argyrous, Statistics for Social & Health Research, op.cit. , p. 465.

شكل (15ـ2) مخرجات SPSS لجداول التقاطع بإدخال متغير التحكم

تفسير مخرجات SPSS لجداول التقاطع مع التحكم في المتغيرات.

إذا نظرنا إلى هذه المخرجات نجد في الحقيقة جدولين متقاطعين جمعا في جدول واحد فالنصف الأول من الجدول، هو جدول التقاطع للدخل مع مشاهدة الإذاعة المرئية للحالات التي لم تتحصل على التعليم العالي. وتحت هذا الجدول مباشرة جدول التوافق لتلك الحالات التي أكملت بعضاً من التعليم العالي. إن نسبة الحالات التي تشاهد مستوىً معيناً من الإذاعة المرئية هي الشيء نفسه لكل فئات الدخل بغض النظر عن المستوى التعليمي. وقد عززت هذه النتيجة بقيم جاما Gamma كما ظهرت في الجدول المُعَنْوَن بـِ Symmetric Measures Table. وهذه القيم هي الشيء نفسه كالتي تم حسابها سابقاً. إن العلاقة بين الدخل ومشاهدة الإذاعة المرئية احتفظت بنفس القوة والاتجاه لكل جدول من الجداول الجزئية[4].

2- العلاقة الكاذبة أو الدخيلة:

نفترض أنه عندما أدخلنا المستوى التعليمي في التحليل تحصلنا بدلاً من الجدول (15-2 و15 -3) على جدول (15-4 و 15-5) من خلال هذين الجدولين (4 و 5) نلاحظ أن العلاقة بين الدخل ومشاهدة الإذاعة المرئية والتي شاهدناها في الجدول الأصلي قد اختفت فجأة. إنه بوضوح يمكننا أن نرى أنه لا يوجد ارتباط يمكن الحديث عنه بين الدخل ومشاهدة الإذاعة المرئية، وما إن أدخلنا مستوى التعليم كمتغير للتحكم حتى إن الارتباط الأصلي بين المتغيرين الدخل ومشاهدة الإذاعة المرئية قد اختفى بإدخال هذا المتغير وأن الطريقة الدقيقة للوصول لهذه النتيجة هي حساب جاما الجزئية من قيم جاما لكل جدول جزئي.

جدول (15 ـ 4) مشاهدة الإذاعة المرئية وفقاً لمستوى الدخل: التحكم في مستوى التعليم (لأولئك الذين لم يتحصلوا على التعليم العالي)

مشاهدة الإذاعة المرئية	الدخل		المجموع
	منخفض	عال	
منخفض	102 % 75	50 % 71	152
عال	34 % 25	20 % 29	54
المجموع	136	70	206
جاما = 0.09 = Gamma			

جدول (15 ـ 5) مشاهدة الإذاعة المرئية وفقاً لمستوى الدخل: التحكم في مستوى التعليم (الذين أكملوا بعضاً من التعليم العالي)

مشاهدة الإذاعة المرئية	الدخل		المجموع
	منخفض	عال	
منخفض	13 % 19	45 % 20	58
عال	54 % 81	184 % 80	238
المجموع	67	229	296
جاما = −0.007 = Gamma			

جاما الجزئية Partial Gamma:

إن قيم جاما التي تم حسابها لكل جدول من الجدولين الجزئيين كانت ذات فائدة لكشف النقاب عن العلاقة الكاذبة أو الدخيلة. ويمكننا أن نرى أن هذه القيم المتعلقة

بجاما ضعيفة من حيث القوى باختلاف قيم جاما المشتركة بالجدول الثنائي الأصلي. في الجدول الأصلي، حيث إن الحالات لم تكن منفصلة وفقاً لمستوى التعليم، فقد وصلت قيمة جاما إلى 0.47. إلا أن قيم جاما لكل جدول من الجداول الجزئية تقرب من الصفر.

إن المنطق وراء حساب جاما الذي تَمَّتْ مناقشته في متن هذا الكتاب المتعلق بالأزواج المتوافقة والأعداد المتعلقة بالأزواج غير المتوافقة. فالأزواج المنسجمة أو المتوافقة كما بَيَّنَّا سابقاً هي مجموعة الحالات الزوجية التي تم ترتيبها بشكل متساوٍ مع كل متغير من هذين المتغيرين، وبالتالي شملت العلاقة الإيجابية بين هذين المتغيرين، بينما على الجانب الآخر، فإن الأزواج المتنافرة من الحالات والتي رتبت بشكل مختلف عن المتغيرين قد اشتملت على علاقة سالبة بين المتغيرين.

إذا ما أضفنا الأزواج المتقاربة أو المنسجمة عبر الجدولين الجزئيين وأضفنا الأزواج المتنافرة عبر نفس الجدولين فإنه باستطاعتنا حساب جاما الجزئية التي تقيس العلاقة المباشرة بين المتغيرين اللذين بدأنا بهما مع إدخال متغير ثالث كمتحكم، قد تم حساب جاما لجمع الأزواج المنسجمة والأزواج غير المنسجمة عبر الجداول الجزئية.

تجدر الإشارة إلى أننا لازلنا نستخدم كل الحالات لتحديد جاما الجزئية، ولكننا الآن نقوم بهذا العمل بعد أن تم فصل الحالات إلى جدولين جزئيين منفصلين. إن عملية حساب جاما الجزئية لهذه البيانات يمكن توضيحها في الجدول (15 – 6).

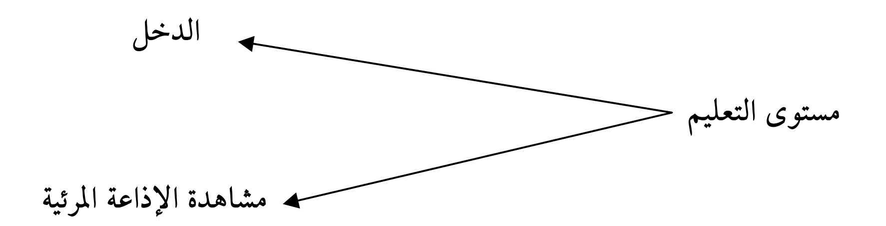
جدول (15 ـ 6) حساب جاما الجزئية

	الأزواج المنسجمة	الأزواج غير المنسجمة	جاما
الجدول الثنائي الأصلي	23.460 = 115 × 204	8360 = 95 × 88	0.47
الجدول الجزئي (1)	2040 = 102 × 20	1700 = 50 × 34	0.09
الجدول الجزئي (2)	2392 = 184 × 13	2430 = 45 × 54	-0.007
المجموع عبر الجداول الجزئية	4432 = 2392 + 2040	4130 = 2430+1700	0.04

إن قيمة جاما الجزئية لهذه البيانات هي فقط 0.04 مشيرة إلى وجود علاقة مباشرة ضئيلة بين الدخل ومشاهدة الإذاعة المرئية عندما تم إضافة متغير مستوى التعليم كمتغير متحكم[5].

العلاقة الكاذبة أو الدخيلة:

عندما كانت جاما الجزئية على درجة منخفضة إذا ما قورنت بقيمة جاما الأصلية التي تم حسابها في جدول التوافق المشترك يتوجب علينا أن نستنتج وجود إما علاقة كاذبة أو دخيلة بين المتغيرين الأصليين. وقبل تفسير كل نمط من أنماط هذه العلاقة، يحتم علينا الإشارة لأن نقرر أياً من النماذج يفسر النتائج المتعلقة بتوسيع الجداول هل هو تفسير نظري وليست مسألة إحصائية وعندما وجدنا أن العلاقة الأصلية قد اختفت بعد توسيع جداول التوافق، فالأمر يبقى لنا لنقرر كيف تتوافق هذه المتغيرات الثلاثة معاً استناداً إلى فهمنا إلى الواقع، فعلى سبيل المثال، يمكننا الاعتقاد أن النموذج الموضح أدناه (شكل 15 -3) هو أفضل تفسير للنتائج التي تم تحليلها الآن.



شكل رقم (15 ـ 3) علاقة كاذبة

إن هذا النموذج يوضح لنا العلاقة الكاذبة بين الدخل ومشاهدة الإذاعة المرئية بإدراك أن هذه العلاقة غير موجودة بين المتغيرين الأصليين وأن أن هذه العلاقة مجرد نتيجة إحصائية فقط استناداً على خصوصية علاقتهما بالمتغير المتحكم. فالمستوى التعليمي يؤثر بشكل منفصل على الدخل ومشاهدة الإذاعة المرئية، إلا أن المتغيرين الأخيرين ليسا مرتبطين ببعضهما البعض.

وقد نجد باحثاً آخر ينظر إلى هذه النتائج التي تحصلنا عليها من خلال توسيع جداول التقاطع بين الدخل ومشاهدة الإذاعة المرئية وبدلاً من ذلك يمكنه وصف العلاقة كما هي موضحة قي الشكل رقم (4).

الدخل ←——— المستوى التعليمي ←——— مشاهدة الإذاعة المرئية

شكل رقم (15 ـ 4) علاقة دخيلة

إن هذا الباحث يمكنه أن يجادل في أن أولئك الذين يكسبون دخلاً عالياً بإمكانهم الحصول على التعليم الجامعي، وهذا التعليم الجامعي يؤثر على كمية مشاهدة الإذاعة المرئية. ومهما كان اعتقادنا بأن هذه المناقشة جيدة أو غير ذلك، فهي مسألة ترجع في الأساس إلى كونها مسألة نظرية. أما إذا كان هذا التفسير أكثر ملاءَمةً لنتائج توسيع الجداول أكثر من نموذج العلاقة الكاذبة فهو أمر متروك لنقاش. لكن التحليل الإحصائي في حد ذاته لا يمكنه أن يقرر بين المسألة. فنتائج التحليل الإحصائي هي مجرد إشارة إلى أن واحداً من هذه النماذج يمكنه أن يفسر النتائج بشكل جيد.

العلاقة المشروطة:

نفترض أن باحثاً ما يرغب في معرفة إلى أي مدى يستجيب المرضى لبرنامج رياضي يهدف إلى تحسين نظام الأوعية الدموية (القلب) وقد قسم الباحث المرضى إلى مجموعتين: مجموعة ذات مستوى منخفض من الممارسة الرياضية والأخرى ذات مستوى عالٍ من الممارسة الرياضية ولاحظ ما إذا كان هناك أي تحسن في أنظمة الأوعية الدموية (انظر جدول رقم 7).

جدول (15 ـ 7) تحسن أنظمة الأوعية الدموية وفقا لمستوى الممارسة الرياضية

التحسن	المستوى الرياضي		المجموع
	منخفض	عال	
لا	38 % 73	11 % 34	49
نعم	14 % 27	21 % 66	35
المجموع	52	32	84

ومن خلال جدول (15 – 7) يمكننا أن نلاحظ أن خلايا النموذج لكل عمود يقترح أن هناك علاقة قوية وموجبة بين المتغيرين. إن البرنامج الرياضي يبدو أنه ذو فعالية. ولتعزيز هذه الفكرة فقد تم حساب جاما التي أنتجت قيمة تساوي 0.68.

إن الباحث يمكنه أن يتوقف عند هذه النتيجة، ويصل إلى قرار أن هناك علاقة مباشرة بين المتغير المستقل (مستوى التمارين الرياضية) والمتغير التابع (مستوى التحسن). وعلى أية حال، فإن الباحث يعتقد أن العلاقة الفعلية أكثر تعقيداً من هذا، إنه ربما توجد عوامل أخرى لم يضعها الباحث في هذا التحليل والتي يمكنها أن تقود إلى ما إذا كان قد تحسن نظام الأوعية الدموية (القلب) لدى المريض.

وبشكل خاص، يمكن للباحث أن يعتقد فيما إذا كان الشخص من المنتظمين في التدخين سوف يؤثر على فرصتهم للاستجابة لبرنامج الممارسة الرياضية. ومن هنا يتوجب على الباحث أن يعيد حساب جاما في جدول التوافق، وإدخاله متغير مستوى التدخين هذه المرة. والنتائج لهذا التحليل تظهر في الجدولين 8 و 9.

جدول (15 ـ 8): التحسن في الأوعية الدموية وفقا لمستوى الممارسة الرياضية (المدخنون فقط)

التحسن	مستوى الرياضة		المجموع
	منخفض	عال	
لا	28 % 74	7 % 70	35
نعم	10 % 26	3 % 30	13
المجموع	38	10	48
جاما = 0.09			

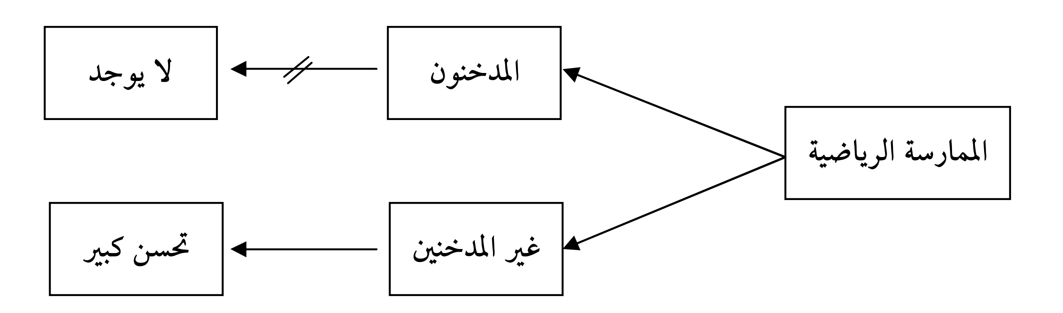
جدول (15 ـ 9) التحسن في نظام الأوعية الدموية وفقا لمستوى الممارسة الرياضية (غير المدخنين)

التحسن	مستوى الرياضة		المجموع
	منخفض	عال	
لا	10 % 71	4 % 18	14
نعم	4 % 29	18 % 82	22
المجموع	14	22	36
جاما = 0.84			

وعند مقارنة هذين الجدولين الجزئيين بالجدول الأصلي فإنه من الواضح أن العلاقة بين هذه المتغيرات تتباين اعتماداً على تاريخ الشخص المدخن. فالأفراد المنتظمين في التدخين لم يسجلوا أي تحسن في مستويات صحتهم كنتيجة لبرنامج الممارسة الرياضية. بينما سجل الأفراد غير المدخنين علاقة أقوى مما كانت عليه عند التعامل مع الجدول الأصلي. فالنتيجة قد ضعفت نتيجة لتضمين أولئك المدخنين مما أدى إلى عدم ثبات العلاقة وهذا يتأكد لنا

من خلال قيم جاما لكل جدول من هذه الجداول. فالجدول المتعلق بغير المدخنين فقط وصلت قيمة جاما إلى 0.84 مقارنة بـِ 0.68 للجدول الأصلي ككل. أما فيما يتعلق بالمدخنين، فإنه لا يوجد أي تحسن من برنامج الممارسة الرياضية، وبالتالي نجد أن النتيجة قد اختفت نتيجة لتأثير المتغير المتحكم. إن مقياس التطابق يعتبر مقيداً للغاية نظراً لأن مقياس التطابق يقيس التغيرات التي حدثت عندما أضيف متغير التحكم.

وكنتيجة لهذه الملاحظة فقد عمد الباحث إلى تغيير النموذج الذي يمكن أن يربط المتغيرات بعضها ببعض. وبدلاً من العلاقة البسيطة المباشرة ذات الاتجاه الواحد، فالباحث يمكنه أن يصف الارتباط في إطار العلاقة المشروطة كما هي مبينة في الشكل التالي:



شكل رقم (15 ـ 5) العلاقة المشروطة

إن العلاقة المشروطة يطلق عليها في بعض الأحيان (التفاعل)، فالتفاعل يكون موجوداً عندما تكون العلاقة بين متغيرين معتمدة على قيم خاصة بمتغير ثالث. وفي بعض الأحيان يمكننا أن نجد العلاقة عكسية اعتماداً على قيمة المتغير المتحكم؛ فقد تكون العلاقة لمجموعة فرعية علاقة موجبة، في حين يمكن أن تكون العلاقة سلبية في مجموعة أخرى فرعية.

مثال:

الآن يمكننا أن نستقصي العلاقة بين مقياس الذكاء (IQ) والدخل. فالذكاء مقياس وفقاً لمعيار اختبار IQ ويمكننا تقسيم المبحوثين إلى مجموعتين: مجموعة ذات مستوى ذكاء

منخفض والأخرى ذات مستوى ذكاء عالٍ. كذلك يمكننا تقسيم هؤلاء المبحوثين أيضاً إلى مجموعتين: مجموعة ذات دخل منخفض ومجموعة ذات دخل عالٍ وذلك طبقاً لمتوسط الدخل القومي لهذا المجتمع أو ذاك؛ إن النتائج المجمعة لـِ: 1000 شخصٍ تم مسحهم يوضحها الجدول رقم (15 – 10)، ويوضح هذا الجدول أن هناك ارتباطاً متوسطاً بين الذكاء – كما تم قياسه باختبار IQ – والدخل، وقد تقود هذه النتيجة إلى تفسير إلى أن التباين في الذكاء يسبب التباين في مستوى الدخل. فقدرة الناس على الكسب تحدد من خلال ذكائهم الخاص. فهم إلى حد ما ومن أجل تحاشي مثل هذه النتيجة يمكننا أن نجادل بأن اختبار (IQ) كقياس للذكاء يكون اختباراً متحيزاً، إننا قد نشعر بشكل خاص أن درجات IQ هي انعكاس لخلفية الوضع الاجتماعي، وأن هذا المتغير هو الأساس في تحديد الدخل ولتقييم هذا الأمر يمكننا بناء جدولين جزئيين من خلال تقسيم الـ 1000 مبحوث إلى فئة اجتماعية عالية وفئة اجتماعية منخفضة، وإن هذه المجموعات الفرعية ولدت لنا النتائج التالية كما توضحها الجداول 15 – 10، 15 – 11، 15 – 12.

جدول (15 ـ 10) الدخل وعلاقته بالذكاء

الذكاء IQ	الدخل		المجموع
	منخفض	عال	
منخفض	165 %36	95 %18	260
عال	295 %64	445 %82	740
المجموع	460	540	1000
جاما = 0.45			

جدول رقم (15 ـ 11) الدخل وعلاقته بالذكاء
(الفئة الاجتماعية العُليا فقط)

الذكاء IQ	الدخل: منخفض	الدخل: عال	المجموع
منخفض	20 %18	60 %14	80
عال	90 %82	380 %86	470
المجموع	110	440	550
جاما = 0.17			

جدول رقم (15 ـ 12) الدخل وعلاقته بالذكاء
(الفئة الاجتماعية الدنيا فقط)

الذكاء IQ	الدخل: منخفض	الدخل: عال	المجموع
منخفض	145 %41	35 %35	180
عال	205 %59	65 %65	270
المجموع	350	100	450
جاما = 0.13			

ومن خلال هذه الجداول يتضح أن هناك علاقةً ثنائيةً قويةً قد اختفت بشكل كبير عندما تم إدخال الخلفية الاجتماعية كمتغير للتحكم. وكما أنه يمكن ملاحظة أنَّ هناك فرقاً بسيطاً في نمط التوزيعات النسبية عبر الجدولين الجزئيين. في حقيقة الأمر أنه عندما تم حساب جاما الجزئية على أساس الجداول الجزئية وجدنا أن قيمة جاما فقط 0.15 في هذه الحالة يمكننا القول بأنّ لدينا علاقة كاذبة أو علاقة دخيلة [6].

الخلاصة:

في هذا الفصل قد تمت مناقشة الطريقة التي من خلالها تم إدخال متغير ثالث يمكنه إحداث تغيير في العلاقة التي تم ملاحظتها سابقاً بين متغيرين. حقاً، قد يكون الأمر أكثر تعقيداً عندما يسمح بوجود تأثير متغيرات أكثر على العلاقة الثنائية الأصلية. آخذين في الاعتبار احتمالية تأثير متغيرات أخرى متعلقة بالتحليل المتعدد، والذي سوف نتناوله في الفصل اللاحق. ولمساعدة القارئ في فهم هذا الفصل يمكننا أن نرسم بعض النتائج من خلال توسعنا في جداول التوافق (انظر جدول 15-13) الذي يقدم لنا دليلاً لاتخاذ القرار.

جدول (15ـ13) النتائج المحتملة عند إدخال متغير التحكم

الجداول الجزئية عند مقارنتها بجداول التقاطع يتبيّن:	النموذج	التطبيقات لتحليل أبعد	احتمالية الخطوة التالية في التحليل الإحصائي	التطبيقات النظرية
العلاقات نفسها بين X و y	علاقة مباشرة	غُض النظر عن متغير التحكم	اختيار متغير تحكم آخر لزيادة اختبار العلاقة المباشرة	نموذج يشير إلى أن متغير X يسبب y في علاقة مباشرة تم دعمها
علاقة ضعيفة أو لا توجد علاقة بين X و y	علاقة كاذبة أو دخيلة	إدخال متغير التحكم أو إدخال متغير متحكم	• التركيز على العلاقة بين هذه المتغيرات الثلاثة • التركيز على العلاقة بين هذه المتغيرات الثلاثة	• النموذج يشير إلى أن X تسبب y لمتغير لم يتم دعمه • النموذج الذي يشير إلى أن X تسبب y جزئياً ولكن يجب تعديل النموذج لأخذ متغير التحكم في الاعتبار
علاقات مختلطة	التفاعل / علاقة مشروطة	إدخال متغير التحكم	تحليل المجموعات الفرعية استناداً على متغير التحكم بشكل منفصل	النموذج الذي يشير إلى أن X تسبب y ثم دعمه جزئياً، ولكن يتوجب علينا تعديل النموذج من أجل أخذ متغير التحكم في الحسبان

المصدر: Joseph F. Healey , Statistics: A Tool for Social Research , Wadsworth Cengage Learning , USA , 1993 , P. 428.

أسئلة للمراجعة:

1- ما هي النتيجة التي يمكنك استنتاجها حول العلاقة بين X و y استناداً على البيانات التالية:

كل الحالات:

Y	X		مجـ
	1	2	
1	177	146	323
2	51	346	397
مجـ	228	492	720

التحكم في C1:

Y	X		مجـ
	1	2	
1	153	52	205
2	44	123	167
مجـ	197	175	372

التحكم في C2:

Y	X		مجـ
	1	2	
1	24	94	118
2	7	223	230
مجـ	31	317	348

2- دراسة أوضحت العلاقة بين العمر، الاهتمام بالبيئة، والانتماء السياسي.. أنتجت لنا قيم جاما التالية:

- جاما (العمر مع الاهتمام بالبيئة): 0.57-
- جاما (العمر مع الاهتمام بالبيئة) (الليبراليين فقط): 0.22-
- جاما (العمر مع الاهتمام بالبيئة) (المحافظين فقط) 0.67-

ما هي النتيجة المستنتجة حول هذه العلاقة؟ إذا كانت هناك علاقة بين هذه المتغيرات.

الهوامش والمصادر:

أولا: الهوامش:

1- George Argyrous , Statistics for Social and Health Research With a Guide to SPSS , Sage Publications , London ,2001 , P.461

2- انظر: عبد الله عامر الهمالي، التقنيات الإحصائية ومناهج البحث، ط 1، منشورات جامعة قاريونس، بنغازي، 2008، ص ص 88-94.

3- George Argyrous , op.cit, PP. 462 - 463.

4- Ibid , PP. 464 - 465.

5- Ibid , PP. 466 - 467.

6- Ibid , PP. 470 - 471.

ثانيا: المصادر:

1- George Argyrous, Statistics for Social and Health Research With Guide to SPSS, Sage Publications, London, 2001.

2- Joseph F. Healey, Statistics for Social Research, Wadsworth Cengage Learning, USA, 1993.

3- ______________ , The Essentials of Statistics: A Tool for Social Research, Wadsworth Cengage Learning, USA, 2010.

الفصل السادس عشر
الارتباط الجزئي والانحدار المتعدد

مقدمة:

نحاول في هذا الفصل أن نتناول بعض التقنيات المستخدمة في تحليل العلاقات السببية التي تساعد الباحث في الوصول إلى تنبُّؤَات حول الظاهرة التي يسعى لدراستها. وتستند هذه التقنيات على ارتباط بيرسون (r) الذي تمتْ مناقشته في الفصل التاسع وتستخدم هذه التقنيات بشكل ملائم وعلى درجة عالية من الدقة، وتقاس متغيرات هذه التقنيات على مستوى المقياس ذي المسافات والنسبي.

إن أول هذه التقنيات التي سوف نناقشها في مستهل هذا الفصل هي تحليل الارتباط الجزئي.

أولاً: الارتباط الجزئي:

تعريفه: الارتباط الجزئي هو عبارة عن تقنيات إحصائية تسمح للباحث أن يفحص العلاقة الثنائية بين متغيرين، بينما يتحكم في المتغير الثالث، والذي يمكن الإشارة إليه في هذا الشأن بـِ (A) أو متغير التحكم.

حساب الارتباط الجزئي:

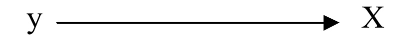
لحساب الارتباط الجزئي - بادئ ذي بدء - ينبغي على الباحث إجراء ارتباط بيرسون (r)، وبعد ذلك يمكنه الدخول في إجراءات حساب الارتباط الجزئي. وتجدر الإشارة هنا إلى الاختلاف بين الارتباط الثنائي (r) والارتباط الجزئي، حيث يفترض الباحث أن هناك تأثيراً للمتغير الثالث على العلاقة الثنائية بين المتغيرين تحت الاستقصاء. فعلى سبيل المثال، إذا كان الزوج المتعلم، والزوج غير المتعلم لديهما استجابات مختلفة فيما يتعلق بعدد الأطفال في الأسرة. فإن الارتباط الجزئي سيختلف في القوة، وربما في الاتجاه إذا قورنا بمعامل الارتباط الثنائي.

وقبل الدخول مباشرة في حساب معامل الارتباط الجزئي، دعنا نأخذ بعين الاعتبار العلاقات بين معامل الارتباط الجزئي ومعامل الارتباط الثنائي، وماذا تعني هذه الارتباطات حيث توجد ثلاثة أنماط محتملة من هذه العلاقات.

أنماط العلاقات[1]:

1- العلاقات المباشرة:

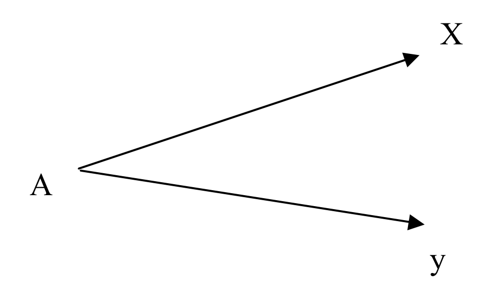
في الحقيقة إنه عندما يُدخل الباحث متغير التحكم في تحليل بياناته، فإنه قد يتحصل على أحد الاحتمالات، وهي أن معامل الارتباط الجزئي في الأساس قيمته مساوية للمعاملات الثنائية. فعلى سبيل المثال، إذا تم التحكم في تعليم الزوج، فإننا نجد أن قيمة معامل الارتباط الجزئي تصل إلى 0.49+ إذا قورنت بارتباط بيرسون (r) 0.50+. وهذا يعني إن إدخال المتغير الثالث لا تأثير له على العلاقة الأصلية (عدد الأطفال وعدد الساعات المتعلقة برعاية هؤلاء الأطفال). بمعنى آخر، أنه بغض النظر عن مستوى التعليم، فإن الأزواج يستجيبون بطريقة واحدة لمسألة رعاية الأطفال. إن مثل هذه النتيجة تكون متوافقة مع النتيجة التي تشير إلى العلاقة المباشرة أو السببية بين X و y.



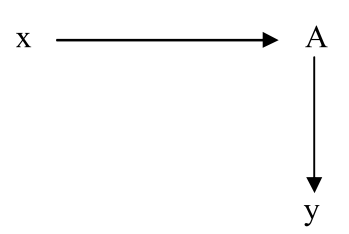
وأن المتغير الثالث لا علاقة له بالاستقصاء. وفي هذه الحالة ينبغي على الباحث أن يغض النظر عن المتغير الثالث (A) في التحليلات الإحصائية اللاحقة.

2- العلاقات الكاذبة والدخيلة:

إن الاحتمال الآخر الذي يمكن أن يحدث هو عندما تكون قيمة معامل الارتباط الجزئي ضعيفة إذا ما قورنت بمعامل الارتباط الثنائي، وربما يصل هذا الارتباط إلى صفر. وقد تكون النتيجة متساوية مع علاقات مختلفة بين المتغيرين. أولهما العلاقة الكاذبة Spurious Relationship. فالتحكم في متغير (A) هو سبب في وجود كلا المتغيرين، المتغير المستقل (X) والمتغير التابع (y)، كما هو موضح في الشكل التالي:



وتعني هذه النتيجة أنه لا توجد علاقة حقيقية بين X و y، ويظهر هذان المتغيريان أنهما مرتبطان بعضهما ببعض فقط بوجود متغير التحكم A. وعندما يضع الباحث في اعتباره متغير التحكم (A)، فإن العلاقة المرئية بين X و y ستختفي. وثانيهما العلاقة الدخيلة بين X و y هذا النمط (يكون الارتباط الجزئي أكثر ضعفاً من معامل الارتباط الثنائي) متسقاً Consistent مع العلاقة الدخيلة بين المتغيرين كما هو موضح في الشكل التالي:



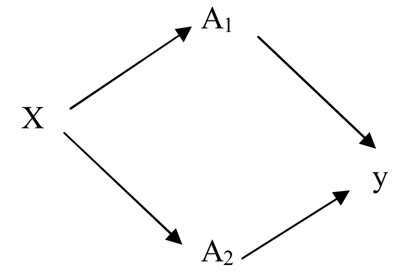
في مثل هذا الموقف، نجد أن X و y متغيران لا يرتبطان بعضهما ببعض مباشرة

ولكنهما – سببياً – مرتبطان بوجود التحكم (A). مرة أخرى، إنه عند التحكم في متغير (A) فإن العلاقة المرئية بين X و y ستختفي.

إن السؤال الذي يمكن طرحه في هذا الخصوص هو: كيف يمكن للباحث أن يميز بين العلاقة الكاذبة والعلاقة الدخيلة؟ للإجابة عن هذا السؤال، يمكننا القول، إن التمييز بينهما لا يمكن القيام به من خلال الأسس الإحصائية: فالعلاقة الكاذبة والعلاقة الدخيلة هما شيء واحد من الناحية الإحصائية: فالباحث يمكنه أن يميز بين هاتين العلاقتين فقط من حيث ترتيب المتغيرات، من حيث الوقت (بمعنى أي من المتغيرين قد جاء أولاً) أو يمكنه التمييز بين هاتين العلاقتين على أسس نظرية *Theoretical grounds* وليس على أسس إحصائية.

التفاعل Interaction:

وأخيراً، فإن الاحتمال الثالث للعلاقة بين المتغيرين يمكن الإشارة إليه هنا، بالرغم من صعوبة اكتشافه من خلال تحليل الارتباط الجزئي، ويطلق على هذه العلاقة: التفاعل. وتحدث علاقة التفاعل بين X و y وتتغير بشكل ملحوظ تحت قيم متنوعة لـِ A. فعلى سبيل المثال، قد تكون هناك علاقة موجبة بين X و y لأحد فئات A، وعلاقة سالبة للفئة الأخرى كما هو مبين في الشكل التالي:



حساب وتفسير معامل الارتباط الجزئي:

المصطلحات والمعادلات: تتطلب معادلة الارتباط الجزئي بعض المصطلحات الجديدة. وسنتعامل هنا مع أكثر من واحد من العلاقات الثنائية، وبالتالي، نحتاج أن نميز بين هذه

العلاقات من خلال رموز معينة. وعليه، فإن الرمز ryx سوف يشير لمعامل الارتباط بين المتغير y والمتغير x، و ryA يشير إلى معامل الارتباط بين متغير y ومتغير A، و rxA يشير إلى معامل الارتباط بين X و A.

معامل الارتباط الجزئي:

إن أول علاقة جزئية يمكن الإشارة إليها بالرموز الجبرية التالية:

ryx.A

يمثل المتغير A إلى اليمين من النقطة، المتغير المتحكم وعليه، ryx.A تشير إلى معامل الارتباط الجزئي التي تقيس العلاقة بين المتغير x والمتغير y مع التحكم في متغير A. والمعادلة المتعلقة بالارتباط الجزئي:

$$ryx.A = \frac{ryx - (ryA)(rxA)}{\sqrt{1 - r^2}\, yA\sqrt{1 - x}A}$$

ملاحظة: ينبغي على الباحث أن يجري أولاً معامل الارتباط بين كل زوجين من المتغيرات (المتغيرات X و y و X و A، و y و X) قبل استخدام هذه المعادلة.

حساب معامل الارتباط الجزئي:

لتوضيح حساب معامل الارتباط الجزئي يمكننا استخدام المثال التالي:

مثال: البيانات التالية تمثل التدخين، الإصابة بالزكام، والإجهاد:

X التدخين	Y الإصابة بالزكام	A الإجهاد
70	2	4
105	6	3
35	1	2
105	4	3
0	2	2
70	7	3
35	3	1
140	6	5
0	4	5
140	4	3
$\overline{X}=70$	$\overline{y}=3.9$	$\overline{A}=3.10$
$S^2X=2450.0$	$S^2y=3.49$	$S^2A=1.49$
$SX=49.50$	$Sy=1.87$	$SA=1.22$

المصدر: George Diekhoff , Statistics for Social and Behavioral Sciences , Univariate , Bivariate , Multivariate , wm. c. Brown Publishers ,USA , 1992 , PP. 257 - 258

rxy =.53

rAy =.40

rAx =.29

$$ryx.A=\frac{ryx-(ryA)(rxA)}{\sqrt{1-r^2\,yA}\sqrt{1-r^2\,xA}}$$

$$=\frac{(.53)-(.40)(.29)}{\sqrt{1-(.40)^2}\sqrt{1-(.29)^2}}$$

$$= \frac{0.414}{\sqrt{0.84}\sqrt{0.92}}$$

$$= \frac{0.414}{(0.92)(0.96)}$$

$$= \frac{0.414}{0.883}$$

$$ryx.A = 0.47$$

التفسير: إن هذه القيمة التي تقيس العلاقة بين التدخين والإصابة بالزكام مع التحكم في متغير الإجهاد هي أقل من معامل الارتباط (r) ryx.53.

إن مربع الارتباط الجزئي هو 22. = $rxy.A = .47^2$. وتشير هذه الدرجة إلى أن 22% من التباين للتعرض للإصابة بالزكام يمكن التنبؤ به من خلال عدد مرات التدخين عندما يتم التحكم في متغير الإجهاد المهني[2].

اختبار الدلالة للارتباط الجزئي:

إن الارتباط الجزئي بين التدخين (X) والإصابة بالزكام (y) والتحكم في الإجهاد المهني (A) تم حسابها rxy.A =.47.

إن السؤال الذي يمكن طرحه هنا هو: ما إذا كانت هذه القيمة للارتباط الجزئي كافية. بمعنى هل هذه القيمة من الارتباط الجزئي دالة إحصائياً؟

دعنا نرى ذلك من خلال اتّباع الخطوات التالية:

الخطوة الأولى: صياغة الفرض الصفري والفرض البديل:

Ho: إن القيمة المشاهدة لِـ rxy.A مرجعها إلى خطأ المعاينة. إن الارتباط الجزئي بهذا الحجم نسبياً يكون مرجحاً أنه يوجد في عينة سحبت من مجتمع يكون فيه قيمة rxy.A تساوي صفراً. $rxy.A = 0$

Hi: إن القيمة المشاهدة لـ rxy.A كبيرة جداً لأن تكون مرجحة لخطأ المعاينة. إن هذه القيمة من غير المحتمل، أن العينة المسحوبة من مجتمع تكون فيه rxy.A تساوي صفراً

وتعطي درجة عالية للارتباط الجزئي. وعليه، فإن احتمال العينة قد جاء من المجتمع الذي تكون فيه 0 < rxy.A، وأن هناك احتماليةً كبيرةً بأنه في حالة إعادة الدراسة فإن درجة الارتباط الجزئي تكون أكبر من صفر (0).

الخطوة الثانية: إحصاء الاختبار The Test Statistic:

إن الارتباط الجزئي rxy.A يساوي 47. يمثل إحصاء الاختبار للعلاقة بين X و y بعد التحكم في متغير A.

الخطوة الثالثة: تحديد الاحتمالية المرتبطة بإحصاء الاختبار.

بمقارنة القيمة المحسوبة لِـ rxy.A المساوية لِـ 47. لتوزيع المعاينة rxy.A نستطيع عندئذ rxy.A تحديد الاحتمالية القريبة بأن العينة التي تعطي هذه القيمة من الارتباط الجزئي قد جاءت من مجتمع فيه تكون قيمة rxy.A تساوي صفراً (0). ولدرجة حرية df=N-J=10-3=7، حيث (N) تشير إلى عدد الحالات و(J) تشير إلى عدد المتغيرات الخاضعة للتحليل، وباعتماد اختبار أحادي الجانب، فإن القيمة الحرجة لِـ (r) كما هو موضح في جدول توزيع قيمة (r) تحت مستوى دلالة 05. تكون القيمة مساوية لِـ rxy.A تساوي 47. تقع تحت القيمة الحرجة، فإننا بالتالي نصل إلى نتيجة مفادها أن الاحتمالية أكبر من 0.05. إن العينة المسحوبة من مجتمع تكون فيه rxy.A تساوي صفراً (0) ستعطينا ارتباطاً جزئياً يصل إلى 47. = rxy.A، وعليه فإن العينة المدروسة إلى حد كبير يرجح أنها تمثل هذا المجتمع. بمعنى آخر، إن العلاقة بين التدخين والإصابة بالزكام ليست دالة إحصائياً عند التحكم في الإجهاد المهني إحصائياً[3].

الارتباط الجزئي باستخدام SPSS:

1- إدخال البيانات Data Entry:

إدخال قيمة X في العمود الأول، وقيم y في العمود الثاني و A في العمود الثالث في محرر بيانات Spss.

2- تحليل البيانات Data Analysis:

1- انقر على Anayze ⟸ واختر Correlate وانقر على Partial.

2- في الوقت نفسه يلقى الضوء على قائمة العمود لِـ X و y على الجانب الأيسر من العمود ويتم تحريكها إلى مربع المتغيرات.

3- أَلْقِ الضوء على قائمة العمود لِـ A ويتم التحرك إلى مربع التحكم.

4- انقر على OK.

3- مخرجات Spss:

تمدنا هذه المخرجات بمصفوفة الارتباط Correlation Matrix، الارتباطات المحتملة بما فيها الارتباط بين X و X، والارتباط بين y و y. نحن نرغب في الارتباط الجزئي بين X و y التي يكون في أعلى زاوية من يمين المصفوفة. وتحت الارتباط يظهر مستوى الدلالة Significance Level. وعندما يكون مستوى الدلالة أقل من 05. فإن ذلك يشير إلى أن معامل الارتباط معامل دالة.

ثانياً: تحليل الانحدار المتعدد:

مراجعة الانحدار الثنائي Bivariate Regression:

في المثال التالي يمكننا أن نستقصي العوامل التي تؤثر في قيمة مبيعات ما. لقد قامت إحدى الوكالات للخدمات العقارية بجمع بيانات حول متغيرين لعدد 12 منزلاً. وقد جاءت النتائج وفقاً للجدول التالي:

جدول (16 ـ 1) حجم المنزل وأسعار البيع

سعر البيع (000،$)	حجم المنزل	سعر البيع (000،$)	حجم المنزل
260	20	287	24
240	18	252	20
245	20	270	23
210	13	275	25
230	18		
242	14		
295	28		
235	16		

المصدر: George Argyrous, Statistics for Social & Health Research, with a Guide to SPSS , SAGE Publications, London ,2001 , p. 475

إن الغرض من هذا التحليل هو تفسير التباين في سعر بيع هذه المنازل التي تم التعامل معها كمتغير تابع. وتعتقد وكالة الخدمات العقارية أن العامل الأساسي المفسر للتباين في أسعار البيع يرتبط بالتباين في مساحات هذه المنازل. وبالتالي يمكن أن نطلق عليه نموذج العوامل المحددة لسعر المنزل. وبما أن هذا النموذج هو نموذجُ نظريٌ يصف العلاقة التي يمكن أو لا يمكن أن تصمد أمام التدقيق الأمبيريقي.

دعنا نقارن، على سبيل المثال، منزلين من عينة هذه المنازل، كالمنزل الذي تم بيعه بقيمة 252,000 ألف دولار، بالمنزل الآخر الذي تم بيعه بقيمة 230,000 ألف دولار. من خلال هذه المقارنة، يمكننا القول، بأن قيمة البيع مرتبطة بحجم المنزل. ومن هنا نجد أن هذين المنزلين يبدو أنهما متناغمان مع نموذج هذه الوكالة للخدمات العقارية.

إن السؤال الذي يمكن طرحه هنا هو: هل هذه العلاقة تبقى صحيحة في كل المنازل الأثنى عشر؟

وبتطبيق تحليل الانحدار البسيط لهذه البيانات الواردة في الجدول أعلاه مستخدمين طريقة أقل المربعات المألوفة Method of Ordinary Least Square، تتولد لدينا النتائج التالية:

$$y = 157 + 4.88\ X$$

$$r = 0.92$$

$$r^2 = 0.85$$

وعلى ضوء هذه النتائج يمكننا التوصل إلى:

- هناك علاقة موجبة بين حجم المنزل وسعر البيع.
- أن أي زيادة في حجم المنزل تقود إلى الزيادة في سعره بحوالي 4880 دولاراً.
- إن هذه العلاقة، علاقة قوية وبدرجة عالية من الثقة لإجراء عملية التنبؤ.
- إن التباين في حجم المنزل ليس بالضرورة أن يكون على نحو كامل للتنبؤ بسعر المنزل، فمعامل التحديد، معامل قيمته عالية (0.85)، ولكنها ليست مساوية لواحد صحيح. ومع ذلك يمكن القول، بأن هناك عوامل أخرى تؤثر في سعر عينة هذه المنازل.

إن ثمن البيع الفعلي لأي منزل من هذه المنازل يمكننا التعبير عنه بالمعادلة التالية:

ثمن البيع = a + b (حجم المنزل) + e

وتشير هذه المعادلة إلى أن ثمن مبيعات هذه المنازل تتباين بشكل أساسي بسبب الفروق في أحجامها، وكذلك بسبب عوامل عشوائية تمثلت من خلال مصطلح الخطأ (e). ويشير مصطلح الخطأ (e) إلى الفرق بين ما تم التنبؤ به فيما يتعلق بثمن المنزل من خلال حجمه المحدد، والثمن الذي بَيْع به فعلاً.

وتجدر الإشارة إلى أن هناك عدة عوامل تؤثر في مبيعات هذه المنازل، إلا أن النموذج المعتمد لدى وكالة الخدمات العقارية ترى أنه بالرغم من العوامل الأخرى التي تؤثر في مبيعات هذه المنازل، إلا أنه من بين هذه العوامل لا يوجد متغير واحد (حجم المنزل) الذي يلعب دوراً أساسياً في تحديد ثمن البيع بطريقة نظامية وثابتة. وهذا ما حدا بنا إلى التركيز على متغير حجم المنزل وإعطائه وضعاً بيّناً في المعادلة. إلا أنه مع ذلك يمكننا القول، أيضاً إننا لا نريد أن نتجاهل العوامل الأخرى المؤثرة في ثمن بيع هذه المنازل.

إن مصطلح الخطأ (e) يخدم كل هذه العوامل الأخرى، فالعوامل التي تؤثر في ثمن بيع المساكن في كيفما اتفق وبطريقة غير نظامية. فقد يباع أحد المنازل بقيمة عالية، لأن وكالة الخدمات العقارية كانت، بشكل خاص، تطرح هذه المنازل للبيع بطريقة مغامرة ووضع ثمن معين لهذا المنزل، في حين يباع منزل آخر، لأن المشتري قد جذبه بشكل خاص نظام الألوان؛ في حين، يمكن أن يباع منزل آخر بثمن بخس لأن البائع Vendor يريد بيع المنزل بشكل سريع نظراً للدين محدد السداد، ونتيجة لهذه العوامل وعوامل أخرى، تنبثق عشوائياً من بيع إلى بيع آخر الذي لا ينبغي علينا أن نتعامل معها كمتغيرات مستقلة ومنفصلة. ولكن بالسماح لمصطلح الخطأ أن ينتزع تأثير هذه العوامل مجتمعة. ففي بعض الأحيان، هذه العوامل العشوائية تُسَبِّبُ زيادةً عاليةً في عملية البيع أكثر مما نتنبأ به استناداً على المعرفة المتعلقة بحجم المنزل، وفي أحيان أخرى تتسببَّ هذه العوامل في انخفاض ثمن البيع بحيث يكون أقل من القيمة المتوقعة. وبمعرفة حجم المنزل سوف يُسْمَح لنا بالتنبؤ بقيمة البيع التي سوف تكون قريبة من الثمن المستهدف. إلا أننا نسلّم بأن منزلاً معيناً فإن تأثير هذه العوامل العشوائية سوف تعني أن ثمن البيع الحقيقي ليس بالضرورة أن يكون مساوياً للثمن الذي تمَّ التّنُّبؤ به[4].

مقدمة للانحدار المتعدد:

تجدر الإشارة إلى أنه يمكن أن ننظر إلى النموذج الثنائي Bivariate Model بشكل مبسط أكثر مما ينبغي. فقد يتولد لدينا شعورُ بأن هناك عواملَ أخرى غيرَ عامل (حجم المنزل) قد لا تكون عواملَ عشوائيةً، ولكنها عواملُ تعملُ بطريقةٍ نظاميةٍ مُسَبِّبَةٍ التباينَ في قيمة مبيعات المنازل وتتباين هذه العوامل بشكل مستقل عن حجم هذه المنازل. بمعنى آخر، إذا قارنا منزلين لديهما نفس الحجم، فالفرق في ثمن البيع لهذين المنزلين قد لا يكون راجعاً فقط للعوامل العشوائية التي ناقشناها للتو. فإذا نظرنا إلى الجدول رقم (1). فإننا نجد أن هناك ثلاثةَ منازلَ في عينة الدراسة - على سبيل المثال - لديها نفس الحجم. فقد تم بيع أحد هذه المنازل بِـ 260,000 ألف دولاراً، والآخر بـِ 245,000 ألف دولاراً، والثالث بـِ 252,000 ألف دولاراً. فالسؤال الذي يمكن طرحه هنا هو لماذا هذه الفروق في أسعار البيع؟ بالرغم من أن هذه المنازل متساوية من حيث الحجم؟ إذا وضعنا ثقتنا

التامة في النموذج الثنائي والقول بأنَّ العوامل العشوائية مسئولة عن تفسير هذه الفروق. أو يمكننا القول، بأنَّ نموذجاً آخرَ يمكن أن يسمح لفعالية متغيرات أخرى بشكل نظامي للتأثير على سعر البيع خَرَجْنَا بتفسيرات أفضل.

ويمكننا، على سبيل المثال، الاعتقاد بأن عمر المنزل أيضاً (وبشكل جزئي) يفسّر هذا التباين في أسعار هذه المنازل. بمعنى، أن عمر المنزل ليس متغيراً أحياناً يمكن أن يؤثر على ثمن بيع المنزل، ولكن بدلاً من ذلك، يعتبر عامل شائع وأنه على نحو منتظم يؤثر على ثمن بيع المنازل.

إن النموذج الجديد يمكن أن يظل قائماً بشكل معقول لتوقع أنه كلما زاد عمر المنزل قل سعره. بمعنى، أننا نتوقع أن هناك علاقة سلبية بين سعر المنزل وعمره. وإذا ما اعترانا الشك فإن ذلك كذلك، فإن الأمر يتطلب منا أن نوسع من تحليل الانحدار ليشمل هذا التحليل العملية المتعلقة بهذا المتغير الآخر، وهذه العملية تشبه إلى حد كبير نفس الطريقة التي اتبعناها في الفصل الخامس عشر المتعلق بالتوسع في التحليل الثنائي لجداول التقاطع والأخذ في الاعتبار التأثير المتوقع للمتغير الثالث. وعند التعامل مع بيانات المقياس ذي المسافات والنسبي كما هو الحال الآن يكون الهدف من وراء تحليل الانحدار المتعدد *Multiple Regression*. فالانحدار المتعدد هو ذلك التحليل الذي يستقصى العلاقة بين مجموعة المتغيرات المستقلة (متغيرين أو أكثر) ومتغير تابع.

واستناداً على هذا النموذج المتعدد، توضح البيانات الواردة في الجدول (16 – 2) عمر الاثنى عشر منزلاً.

جدول (16 ـ 2) ثمن البيع، حجم المنزل، عمر المنزل (12 منزلاً)

ثمن البيع ($0.000)	حجم المنزل	عمر المنزل (بالسنوات)
260	20	5
240	15	12
245	20	9
210	13	15
230	18	9
242	14	7
295	28	1
235	16	12
287	24	2
252	20	5
270	23	5
275	25	5

المصدر: George Argyrous, Statistics for Social & Health Research, op.cit , p. 478

بشكل عام، يمكننا التعبير عن العلاقة بين أي متغير تابع، وأي عدد من المتغيرات المستقلة (K) بالطريقة التالية:

$$y = a + b_1 x_1 + b_2 X_2 + ... b_k x_k + e$$

ومن خلال المثال الذي بين أيدينا، فإنه يمكننا أن نوضح نموذج العلاقة بالتعبيرات التالية:

(ثمن البيع) a + b_1 (حجم المنزل) + b_2 (العمر) + e

بمعنى آخر، إننا نعتقد بأن ثمن بيع المنزل ينطلق في اتجاه واحد أو آخر وفقاً لعمر وحجم المنزل. وإننا بذلك نتوقع أن المنزل القديم ذا الحجم الصغير نسبياً سيئُول سعره نحو الاتجاه الأسفل عبر العملية المستقلة لكلا المتغيرين العمر والحجم. وبشكل عكسي، فإننا نتوقع أن منزلاً حديثاً وكبير الحجم نسبياً سيتجه سعر بيعه نحو الأعلى. وفي شواهد أخرى، فإن حجم وعمر المنزل من الممكن أن يتجها في اتجاهات معاكسة.

وبالرجوع إلى المعادلة السابقة. فإننا لازلنا نسمح للعوامل العشوائية أن يكون لها تأثير. وبالتالي فإن عمر وسعر المنزل لا يحددان تماماً ثمن البيع في كل حالة. ولكن إذا ما كان هذا النموذج المتعدد Multivariate Model هو الأفضل الذي يمكن الركون إليه في تفسير قيمة بيع المنازل، إذا ما قورن بالنموذج الثنائي Bivariate Model، فإن كمية التباين قد تركت لكي تفسر من خلال مصطلح الخطأ (e)، ستكون أصغر في النموذج المتعدد إذا ما قورنت بالنموذج الثنائي الذي بدأنا به في مستهل هذا الفصل. إلا أنه على الجانب الآخر، فإن إدخال العمر في المعادلة لم يكن قادراً على التقليل من نسبة التباين في ثمن البيع الذي يعزى إلى الخطأ (e). إذاً من خلال معرفتنا بعمر المنزل، ليس بالضرورة، أن يُحسِّن من قدرتنا بالتنبؤ بسعر بيعه. ومن هنا نكون قد أضعنا الوقت والجهد في جمع معلومات حول متغير لا فائدة ترجى منه.

إن مهمة الانحدار المتعدد هي أن يوزع التباين في سعر المنازل لكل واحد من هذه العوامل المتنافسة في التأثير على المتغير التابع، هل يهيمن أحد هذه العوامل على تحديد سعر البيع، كقولنا إن عمر أو حجم المنزل يكون بشكل واضح أكثر أهمية. أم أن هذين العاملين يكونا لديهما نفس التأثير؟

والسؤال الآخر الذي يمكن طرحه في هذا السياق، ما الدور الباقي للعوامل العشوائية؟ إن التحليل المتعدد، وفقاً لحساب معامل الانحدار أو الارتباط الجزئي لكل متغير يمدنا بمقاييس دقيقة للتأثير الخاص لهذه المتغيرات المستقلة على المتغير التابع. إنه من الممكن استخدام تقنيات عددية لحساب معامل الانحدار بين كل واحد من هذه المتغيرات المستقلة والمتغير التابع. ومع ذلك، فإن هذه التقنيات ستكون مُرْهِقَةً بشكل كبير ومُسْتَهْلِكَةً للوقت، خاصة إذا كان الباحث يتعامل مع مجموعة كبيرة من البيانات. وتجدر الإشارة هنا إلى أنه، ومع تقدم تقنيات الحاسوب، أصبح من السهولة بمكان على الباحث حساب الانحدار المتعدد من خلال البرامج المعدة لذلك مثل SPSS الذي يقوم بإنهاء العمليات الحسابية بشكل سريع ودقيق. وتبقى مهمة الباحث الأساسية هي تفسير هذه النتائج[5].

إجراءات الانحدار المتعدد باستخدام SPSS:

إن إجراء حساب الانحدار المتعدد يشبه إجراء حساب الانحدار الثنائي الذي تناولناه في الفصل التاسع من هذا الكتاب، ما عدا أننا في تحليل الانحدار المتعدد نقوم بلصق Paste أكثر من متغير إلى المتغيرات المستقلة Independent (s) إلى القائمة المستهدفة Target List.

الإجراء:

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Analyze ⟸ Regression ⟸ Linear

(يعطيك صندوق الحوار لِـ Linear Regression Dialog box)

2- انقر على Selling Price (سعر البيع) في قائمة المتغير Variable List.

3- أنقر على ◄ الذي يشير إلى المتغير التابع Dependent: القائمة المحددة للمتغير. هذه العملية ستقود إلى لصق (Pastes) Selling Price كمتغير تابع.

4- انقر على House Size (حجم المنزل) في قائمة المتغير، وبينما تضغط على مفتاح Shift، انقر على العمر بالسنوات Age in years.

5- انقر على ◄ الذي يشير إلى المتغير المستقل (المتغيرات) Independent (s) في القائمة المحددة للمتغيرات. هذه العملية ستقود إلى عملية لصق كلا المتغيرين House Size وعمر المنزل بالسنوات Age in years كمتغيرين مستقلين.

6- انقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

Variables Entered / Removed[b]

Model	Variables Entered	Variables Removed	Method
I	Age in years House size (Squares[a])		ENTER

a. All Requested Variables Entered

b. Dependent Variable: Selling Price ($0.000)

Model Summary

Model	R	R Square	Adjusted R Square	Std Error of the Estimate
I	.959[a]	.919	.901	.807

a. Predictor: (Constant) , age in years , House size (Squares)

Anova[b]

Model	Sum of Squares	Df	Mean Square	F	Sig.
I Regression Residual Total	.7596248 .158548 .9176796	2 9 11	24.37931 .90660	.29851	.000

a. Predictors (Constant) , age in years , House size (in Squares)

b. Dependent Variable: Selling Price ($0.000)

Coeffi cients[a]

Model	Unstandarized coefficients		Standariz ed coeffi cients	T	Sig.
	B	Std Error	Beta		
I (Constant)	.290224	.22226		.5538	.000
House size (Squares)	.5782	.973	.487	.6502	.026
Age in years	-2.974	.0761	-.503	-2.764	.022

a. Dependent Variable: Selling Price ($0.000)

المصدر: George Argyrous, Statistics for Social & Health Research, op.cit, p. 481.

شكل رقم (16 ـ 1) مخرجات SPSS للانحدار الخطي

إذا نظرنا إلى هذه المخرجات، فإننا نجد كثيراً من المعلومات قد تولدت من خلال برنامج SPSS، وعليه، ينبغي على الباحث أن يركز على أهم الأجزاء في هذه المخرجات. فالجدول المُعَنْوَنُ بِـ Variables Entered / Removed، متغيرات أدخلت / أبعدت، يبين لنا وصفاً لفظياً بسيطاً للنماذج التي تمكننا من التكهن.

إن برنامج SPSS، بإمكانه أن يقوم بحساب عدة انحدارات متعددة بشكل تلقائي، وذلك من خلال استخدام عدة مجموعات من المتغيرات المستقلة ليحدد أيًّاً من هذه المجموعات أو التوليفات تعطينا أفضل تفسير للتباين في المتغير التابع. هنا في هذا المثال، تنبأنا بنموذج واحد أطلق عليه نموذج (1)، الذي استخدم فيه متغير العمر بالسنوات وحجم المنزل بالمربعات كمتنبئ للمتغير التابع قيمة البيع (000.$).

أما الجدول الثاني، والذي أطلق عليه خلاصة النموذج Model Summary فإنه يبين لنا قيمة معامل الارتباط التي تشير إلى قوة العلاقة بين مجموعة المتغيرات المتوالفة في النموذج Model، والمتغير التابع. وتشير قيمة R إلى 959.، علاقة قوية. أن قيمة R هي الارتباط المتعدد مرادفاً لمعامل الارتباط الثنائي، (r). وأن R^2 (معامل الانحدار المتعدد للتحديد) تصل قيمتها إلى 919. وعندما يتم استخدام النموذج الثنائي Bivariate Model لتفسير سعر البيع (عندما تجرى عملية الارتباط فقط مع حجم المـنزل House Size) فإن قيمة معامل التحديد قد وصلت إلى 0.85. وعندما نستخدم كلا المتغيرين: حجم المنزل وعمره للتنبؤ بسعر البيع، فإن معامل التحديد ترتفع لتصل إلى 919.. وتشير هذه القيمة إلى قدرتنا على التفسير (أو التنبؤ) بقيمة بيع المنزل التي زادت عندما توفرت لدينا معلومات حول عمر وحجم المنزل. إن جزءاً من التباين في قيمة البيع التي أرجعناها سابقاً إلى عوامل عشوائية، هي في حقيقة الأمر تعزو للتأثير النظامي لعمر المنزل.

أما الجدول المُعَنْوَن بِـ ANOVA أنوفا فهو جدولُ يحتوي على إحصائيات استدلالية inferential Statistics، تسمح لنا بإجراء استدلالات من خلال العينة على المنازل الأثنى عشرة. إلى هذه النقطة فإن التركيز يدور حول الإحصاءات الوصفية للعينة، وعليه، ينبغي على الباحث أن يتجاوز هذا الجزء من هذه المخرجات. والنظر إلى الجدول المُعَنْوَنِ بِـ Coefficients الذي يوفر عناصر معادلة الانحدار Elements of the regression equation

التي تم التنبؤ بها والتي يمكن كتابتها بالشكل التالي:

(حجم المنزل بالمربعات) 2.578 + 224.29 = قيمة البيع (‎$.000‎)

(عمر المنزل بالسنوات) 2.974

وعند قراءة معادلة الانحدار، فإنه من الأهمية بمكان، أن نضع في أذهاننا وحدات القياس التي على ضوئها تم قياس المتغيرات. ففي المثال الذي بين أيدينا أن أي زيادة في كل مربع واحد في حجم المنزل تؤدي إلى الزيادة في سعر المنزل بقيمة تصل إلى 2578 دولاراً. وبالتحرر من هذه العلاقة، فقد وجدنا أيضاً أن زيادة كل سنة في عمر المنزل تقلل من قيمة البيع (لاحظ العلاقة السالبة) على نحو 2974 دولاراً.

دعنا الآن نتعامل مع شيء عملي. نفترض أننا نود أن نشتري منزلاً حجمه 15 متراً مربعاً وعمره خمس سنوات. ما هي توقعاتنا بالقيمة المطلوبة لبيعه؟ نضع المعلومات في معادلة الانحدار:

(5) 2.974 − (15) 2.578 + 224.29 = قيمة البيع (‎$.000‎)

$$= 224.29 + 38.67 - 14.87$$

$$= 248.09$$

بطبيعة الحال، فإننا لا نتوقع أن يكون 248,090دولاراً الثمن الحقيقي الذي يتحقق عندما يتم بيع المنزل فعلياً، حيث إن العوامل العشوائية لازالت تلعب دوراً كبيراً في هذا الشأن. ولكن، وبوجود درجة عالية لمعامل التحديد، فإن هذه العوامل العشوائية ينبغي ألا تكون علة الثمن الفعلي للانحراف كثيراً عن القيمة المتوقعة.

إنه من الصعوبة بمكان استخدام معامل الانحدار لنقيّم الأهمية النسبية لكل متغير مستقل في تحديد قيمة المتغير التابع طالما أن كل متغير مستقل قد تم قياسه بوحدات مختلفة (أحدهما تم قياسه بعدد السنوات، والآخر بالمربعات). وإذا قمنا بقياس حجم المنزل بوحدة أخرى كالقدم المربع، حينئذٍ ستكون قيمة معامل الانحدار لهذا المتغير مختلفة وذلك

لاختلاف وحدة القياس التي تم اعتمادها في قياس هذا المتغير. بمعنى آخر، لا يمكننا القول بأن ذلك بسبب معامل حجم المنزل التي وصلت إلى 2.578. في حين أن قيمة المعامل المتعلقة بالعمر، قد وصلت إلى 2.974-. من هنا نجد أن متغير العمر متغيرُ قوي وله تأثير فاعل على قيمة البيع. وعمود المعامل المعيارية Standardized Coefficients توضح ذلك. وبدون الغوص في تفاصيل كيف تم حساب هذه المعامل المعيارية (يطلق عليها أيضاً أوزان بيتا beta - weights ، فإننا ببساطة نلاحظ أن هذه الأوزان (المعامل) قد أبعدت تأثير وحدات القياس. إنه باستطاعتنا أن نرى أن العمر (508.-) لديه قوة تأثير طفيفة على قيمة البيع، أكثر من التأثير الذي يحدثه حجم المنزل (487.).

الجدول التالي يلخص الدور الذي تلعبه المقاييس المختلفة التي يولدها برنامج SPss:

جدول رقم (16 ـ 3) تفسير مخرجات SPSS

معامل الانحدار ← (Regression Coefficients)	تسمح لنا بإجراء التنبؤ على المتغير المعياري (التابع) استناداً على قيم المتغيرات المستقلة في إطار الوحدات الأصلية للقياس.
المعامل المعيارية ← (Standardized Coefficients)	تسمح لنا أن نميز بين الأهمية النسبية لكل متغير في تحديد قيمة المتغير التابع.
R ←	تشير إلى قوة العلاقة بين مجموعة المتغيرات المستقلة (كتلة واحدة) والمتغير التابع (المعياري).
R^2 ←	تشير إلى كمية التباين في المتغير الذي تم تفسيره خلال مجموعة المتغيرات المستقلة في النموذج (1)، بذلك إشارة إلى ما إذا كان النموذج (1) نموذجاً جيداً للتنبؤ بالمتغير المعياري.

المصدر: George Argyrous, Statistics for Social & Health Research, op.cit, p. 482.

اختبار الدلالة للنموذج المتعدد:

إن المعلومات الحرجة لاختبار الاستدلال يوضحه الجدول المعنون بـِ ANOVA. ويقوم برنامج SPSS بإجراء اختبار F على النموذج بكامله لاختبار الفرضية المتعلقة بمعامل الارتباط Correlation Coefficients لكل المتغيرات التي يحويها النموذج Model. وهي صفر (0). في هذا المثال الذي بين أيدينا، فإن إحصاء اختبار F لهذا النموذج يكون مستوى الدلالة فيه 0.000 وهذا بين لنا أنه على الأقل أن أحد الارتباطات بين كل واحد من المتغيرات المستقلة والمتغير المعياري ليست مساوية لصفر في المجتمع. وتؤكد هذه النتيجة ما تم الحصول عليه في جدول معامل (Coefficients). حيث إن قيمة t الإحصائية لكل متغير مستقل أنها ذات دلالة على مستوى 0.05.

وعليه، فإنه باستخدامنا لاختبار F (F-test) نرى ما إذا كان على الأقل بعض المتغيرات المستقلة في النموذج هي دالة، وأن إحصاءات t (t-Stataitics) لكل متغير على حدة يشير إلى أي من هذه المتغيرات دالة إحصائياً[6].

الانحدار التدرجي Stepwise Regression:

قد يلاحظ مكتب الخدمات العقارية أن هناك متغيرات أخرى ذات أهمية كتلك المتغيرات التي تعاملنا معها، تلعب هي الأخرى دوراً أساسياً في تحديد قيمة بيع تلك المنازل في هذه المنطقة. وبالرغم من القوة التفسيرية للنموذج ذي المتغيرين المستقلين، فإن مكتب الخدمات العقارية يمكنه أن يجادل في أن قدرتنا في التنبؤ بأسعار هذه المنازل ستكون أفضل عند إدخال متغير مساحة الأرض المقام عليها المنزل، كمتغير آخر مستقل. وعليه، فإن مكتب الخدمات العقارية سيقوم بجمع معلومات إضافية حول مساحة الأرض المقامة عليها هذه المنازل الأثنا عشر. والجدول التالي يوضح هذه البيانات الإضافية.

جدول (16 ـ 4) سعر البيع، حجم المنزل، عمر المنزل
والمساحة المقام عليها المنزل (12 منزلاً)

سعر البيع ($0.000)	حجم المنزل (مربعات)	عمر المنزل (بالسنوات)	مساحة الأرض المقام عليها المنزل (بالمتر المربع)
260	20	5	420
240	15	12	640
245	20	9	600
210	13	15	590
230	18	9	700
242	14	7	720
295	28	1	624
235	16	12	590
287	24	2	710
252	20	5	630
270	23	5	700
275	25	5	710

المصدر: George Argyrous, op.cit , p. 483.

الآن، لدينا ثلاثة نماذج لتفسير سعر بيع هذه المنازل. إن أول هذه النماذج: النموذج الثنائي الذي أخذ في اعتباره حجم المنزل، وعلاقته بسعر البيع. وثاني هذه النماذج النموذج المتعدد الذي أخذ في اعتباره حجم وعمر المنزل وعلاقتهما بسعر البيع. في حين يركز النموذج المتعدد الثالث على متغيرات حجم المساحة كمتغير تفسيري:

نموذج (1) سعر البيع = $a + b_1$ (حجم المنزل) + e

نموذج (2) سعر البيع = $a + b_1$ (حجم المنزل) + b_2 (عمر المنزل) + e

نموذج (3) سعر البيع = $a + b_1$ (حجم المنزل) + b_2 (عمر المنزل) + b_3 (مساحة الأرض) + e

إننا لِلتّوِّ قَدْ ناقشنا الطريقة التي من خلالها يمكن الحكم ما إذا كان متغير ما يضيف إلى القوة التفسيرية للنموذج من خلال النظر إلى تأثيره الضمني على قيمة مربع R. فإذا زادت قيمة R تربيع بشكل دالٍ عندما أضيف متغير إلى النموذج. حينئذٍ فالمعلومات الإضافية التي وفرها لنا هذا المتغيِّر المضاف زادت من قدرة النموذج في تفسير التباين في سعر البيع.

تجدر الإشارة إلى أن هناك طريقةً واحدةً لنقرر من خلالها بين عدة نماذج، وهي الشروع في إجراء انحدارات خطية منفصلة استناداً على مجموعة مؤتلفة من المتغيرات المستقلة التي نود أن يشملها النموذج. وبعد ذلك، يمكننا أن نقارن قيم R تربيع لنرى إلى أي مدى، تكون أيٌّ من قدرتنا قادرةً على تفسير التباين في سعر البيع الذي يصل إلى مداه الأعلى عن طريق كل واحد من هذه المجموعة المؤتلفة من المتغيرات المستقلة. على سبيل المثال، إذا قمنا بإجراء الانحدار المتعدد مضافاً إلى هذه الأجزاء مساحة الأرض، فإن قيمة R تربيع ستصل إلى 0.922 وهي قيمة مساوية للقيمة التي تحصلنا عليها في النموذج (2). الذي يحتوي على متغيرين فقط هما عمر المنزل وحجمه. بمعنى آخر، إن مساحة الأرض المقام عليها المنزل لا تؤدي إلى زيادة قدرتنا في تفسير سعر البيع، فالوقت والجهد في قياس هذا المتغير، هما في واقع الأمر عمل ضائع[7].

إن المشكلة الكامنة في هذه الطريقة، أنها طريقة مملة في القيام بإجراء انحدارات منفصلة لكل واحد من هذه النماذج الممكن بناؤها. إضافة إلى ذلك، صعوبة الحكم على مقدار الزيادة في R تربيع لتبرير ما يتضمنه أي متغير في النموذج. إلا أنه من حسن الطالع أن برنامج SPSS قادر على أن يمدنا بطريقة يطلق عليها الانحدار التدرجي حيث تسمح لنا هذه الطريقة بتحديد أي من المجموعة المتآلفة من المتغيرات المستقلة تقدم لنا أفضل تفسير للمتغير المعياري. وتتم هذه العملية بإضافة وحذف متغيرات من العملية الحسابية طبقاً إلى ما إذا كان كل واحد من هذه المتغيرات يحدث دلالة إحصائية تكون سبباً في تغير قيمة R تربيع[8].

إجراءات الانحدار المتعدد التدرجي باستخدام SPSS:

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Analyze ⇐ Regression ⇐ Linear

(تقود هذه العملية إلى مربع الحوار (Linear Regression)

2- انقر على Selling Price (سعر البيع) في قائمة المتغير.

3- انقر على ◄ التي تشير إلى Dependent في القائمة المحددة للمتغير (هذه العملية تقود إلى لصق Selling Price (Pasts) (كمتغير تابع).

4- انقر على House Size (حجم المنزل) في قائمة المتغير، مع الإبقاء ضاغطاً على مفتاح Shift والنقر على Age in years (العمر بالسنوات) وبعدها انقر على Land Size (مساحة الأرض).. (هذه العملية تلقي الضوء على House Size , Age in years and Land Size).

5- انقر على الـ ◄ التي تشير إلى المتغيرات المستقلة Independent (s) في القائمة المحددة للمتغيرات. (هذه العملية تقود إلى لصق Pasts، House Size , Age in years , Land Size كمتغيرات مستقلة).

6- انقر على ◄ القريبة من Enter (تظهر قائمة Adrop - down menu).

7- في القائمة drop-down menu انقر على Stepwise (تقود هذه العملية إلى اختيار الانحدار التدرجي Stepwise regression كطريقة لتضمن وتبعد المتغيرات من الانحدار.

8- انقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء :

Regression

Variables Entered / Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Age in years		Stepwise (criteria) Probability-of-f-to Enter=.050 Probability-of-f-to remove=1.00
2	House size (Squares)		Stepwise (criteria) Probability-of-f-to Enter=.050 Probability-of-f-to remove=.100

a. Dependent Variable: Selling Price ($0.000)

M0del Summary

Model	R	R Square	Adjusted R Square	Std Error of the Estimate
1	.925a	.856	.842	.999
2	.959b	.919	.901	.807

a. Predictors: (Constant) , age in years.
b. Predictors: (Constant) , age in years , House size (Squares).

ANOVA[b]

Model	Sum of Square	Df	Mean Square	F	Sig
1 Regression	.9995820	1	.9995820	.84859	.000 a
Residua	5.91897	10	.59297		
Total	6.917879	11			
2 Regression	.7596248	2	24.37831	.29851	.0006 b
Residua	.158548	9	.90880		
Total	.9176796	11			

a. Predictors (Constant) , age in years.
b. Predictors (Constant) , age in years , House size (Squares)
c. Dependent Variable: Selling Price ($0.000)

Coeffi cients[a]

Model	Unstandarized coefficients		Standarized coefficients	T	Sig.
	B	Std Error	Beta		
1 (Constant)	.702292	.8325		.19350	.000
Age in years	-5.419	702	-.925	-7.723	.000
2 (Constant)	.290224	.22226		.5638	.000
Age in years	-2.974	.0761	-.509	-2.764	.022
House size (Squares)	.5782	397	.487	.6502	.026

a. Dependent Variable: Selling Price ($0.000)

Excluded Variables[a]

Model	Beta IN	t	Sig.	Partial correlation	Collinearit Statistics
					Tolorance
1 House size (Squares)	.487a	.6502	.026	.662	.265
Land size in meters squares	.015	.115	.911	.038	.973
2 Land size in meters squares	.015b	.147	.886	.052	.973

a. Predictors in the model (Constant) , age in years.
b. Predictors in the model (Constant) , age in years , House size (Squares)
c. Dependent Variable: Selling Price ($0.000)

المصدر: George Argyrous, Statistics for Social & Health Research, op.cit , p. 486.

شكل (16 ـ 2): مخرجات SPSS للانحدار الخطي

ولتفسير هذه المخرجات:

الجدول الأول المُعَنْوَن بِـ متغيرات / أدخلت وأبعدت Variables Entered / Removed في هذا الجدول تم توليد نموذجين من ثلاثة متغيرات تم افتراضها: الأول متضمناً متغير العمر بالسنوات (النموذج الثنائي الأساسي) والذي أطلق عليه في برنامج SPSS ، نموذج (1) (Model 1)، والنموذج الآخر يشتمل على متغير العمر وحجم المنزل (بالمربعات)، والذي أطلق عليه في برنامج SPSS نموذج (2) (Model 2)، أما باقي المخرجات فهي أساساً نفس المخرجات السابقة التي تم توليدها بشكل منفصل من قِبل كل واحد من هذه النماذج، وهنا لدينا نموذجان تم تقديمهما بنفس التحليل. أما الجزء الجديد من المخرجات فيوضحه الجدول الأخير المُعَنْوَنُ بِـ "Excluded Variables"، المتغيرات المستثناه. ويوضح هذا الجدول، وعلى أساس اختبار F وعلى أساس التغيرات في R^2. إن مساحة الأرض المقام عليها المنزل ليست متغيراً مهماً يمكن أن يضمن في أي من هذه النماذج Models. ويعني هذا كله، أن تعقيد النموذج بإضافة هذا المتغير الجديد لن يُؤَمن لنا أي نوع من الدقة فيما يتعلق بتقدير المتغير التابع، وبالتالي ينبغي الاستغناء عنه(9).

الافتراضات التي ينبغي مراعاتها عند استخدام الانحدار المتعدد:

على الرغم من قوة الانحدار المتعدد كأداة قوية لتقييم تأثير مجموعة من المتغيرات المستقلة على المتغير المعياري، إلا أن هناك مجموعةً من الافتراضات حول هذا التحليل تقلل من تطبيقاته. إن كل الافتراضات التي تمت مناقشتها عند الحديث عن الانحدار الثنائي فهي قائمة في تطبيقها على حالة الانحدار المتعدد. إضافة إلى هذه القائمة فإننا بحاجة إلى إضافة افتراض آخر مهم وهو أن التحليل المتعدد يفترض أن كل متغير من المتغيرات المستقلة مستقل عن الآخر. أي تسامت متعدد Mutticollinearity (أي التعدد الذي يقع على نفس الخط).

في المثال الذي بين أيدينا يمكن توضيحه بالشكل التالي:

إن كلاً من عمر وحجم المنزل يؤثران في سعر بيعه، ولكنهما لا يؤثران في بعضهما البعض. وقد يبدو لنا أن هذا الأمر معقول لهذه المتغيرات تحديداً: فإذا تم بشكل مفاجئ توسيع المنزل، فإن هذا الأمر لا يقود بشكل مفاجئ لأنْ يجعل من هذا المنزل منزلاً عتيقاً أو جديداً! وبشكل مشابه إذا ما زاد عمر المنزل، لا يعني بالضرورة أن يزيد حجمه أو يصغر. إن هذه الفرضية تحتوي ضمناً الانحدار المتعدد، مما يجعله إلى حد ما أكثر تعقيداً من التقنيات المتعددة الأخرى التي تناولناها في الفصل السابق، حيث تم استخدام التحليل المتعدد، لتحديد أي من هذه النماذج يقدم لنا أفضل تفسير للعلاقة بين المتغيرات الثلاثة أو أكثر منها[10]. ومن خلال تحليل الانحدار، فإننا نفترض نموذجاً محدداً كما تم في مخرجات SPSS شكل رقم (1)[11].

حدود استخدام الانحدار المتعدد والارتباط:

يعتبر الارتباط الجزئي، والارتباط المتعدد، تقنيات إحصائية قوية لتحليل العلاقات المتبادلة بين ثلاثة أو أكثر من المتغيرات. وتسمح هذه التقنيات للباحث بالقدرة على التنبؤ بدرجات متغير واحد من اثنين أو أكثر من متغيرات أخرى.

ولما كانت هذه التقنيات على درجة عالية من القوة فهي بالتالي تتطلب جهداً كبيراً، ودرجة عالية من البيانات، وقياساً على المستويين ذي المسافات والنسبي، وهذه متطلبات، عادة يصعب على الباحث إنجازها. إضافة إلى ذلك، فإن هذه التقنيات تستند على افتراضات: أن العلاقة المتبادلة بين المتغيرات تتبع شكلاً محدداً.

أولاً، تفترض هذه التقنيات أن كل متغير مستقل لديه صبغة العلاقة الخطية مع المتغير التابع. ولكي يتعرف الباحث على هذه العلاقة الخطية عليه القيام بمراجعة لشكل الانتشار.

ثانياً: تفترض هذه التقنيات أن تأثير المتغيرات المستقلة هي تأثيرات إضافية additive. وهذا يعني، أنه ينبغي علينا أن نفترض أفضل تنبؤ للمتغير التابع (y) يمكن الحصول عليه، وذلك من خلال إضافة درجات المتغير المستقل بعضِها إلى بعضٍ، على سبيل المثال (الفقر عندما يوجد مع الانعزالية) يولد درجة عالية من معدلات الجريمة.

ثالثاً: أن تقنيات الانحدار المتعدد، ومعامل الارتباط تفترض أن المتغيرات المستقلة متغيرات غير مرتبطة. هذه الحالة تعني أن معامل الارتباط بين كل أزواج المتغيرات المستقلة ينبغي أن تكون صفراً. في الممارسة العملية، ومع ذلك نعمل كما لو أن هذا الافتراض قد تم تحقيقه، إذا كانت الارتباطات المتبادلة بين المتغيرات المستقلة على درجة منخفضة.

وأخيراً، يجب الإشارة إلى أننا في هذا الفصل قد غطينا فقط التطبيقات البسيطة للارتباط الجزئي، والانحدار المتعدد، من حيث المنطق والتفسير، ولزيادة التفصيل لهذه التقنيات عندما يكون لدينا متغيرات مستقلة كثيرة واضحة المعالم نسبياً، ومع ذلك، فإن حسابات مثل هذه المواقف تعتبر معقدة جداً. وعندما يواجه الباحث مثل هذه المواقف لمتغيرات تزيد عن ثلاثة حينئذٍ يتطلب منه الاستعانة ببرنامج SPSS أو SAS للتعامل مع هذا الكم الهائل من المتغيرات[12].

أسئلة للمراجعة:

1- من البيانات التالية:

المدينة	الاحتجاج (y)	معدل البطالة (x)	% الدعاية السلبية
A	55	5	60
B	60	8	63
C	65	9	55
D	68	9	53
E	70	10	44
المتوسط	63.6	8.2	$\bar{x}$ = 55.8
الانحراف	5.5	1.7	S = 5.3
الاحتجاج	0.95 معدل البطالة		% الدعاية السلبية −0.87
معدل البطالة	−		−0.70

المطلوب:

a) حساب الارتباط الجزئي بين y و x، والتحكم في A ما هو التأثير الذي يحدثه متغير التحكم على العلاقة الثنائية. هل العلاقة بين الاحتجاج ومعدلات البطالة علاقة مباشرة.

b) احسب معامل الارتباط الجزئي بين (y) الاحتجاج، و (x) الدعاية السلبية متحكماً في معدل البطالة (A). ما هو التأثير الذي يحدثه متغير التحكم على العلاقة الثنائية. هل العلاقة بين الاحتجاج والدعاية السلبية علاقة مباشرة؟

c) اكتب ملخصاً قصيراً يبين نتائج هذه العلاقات بين المتغيرات الثلاثة.

2- من البيانات التالية: احسب معامل الارتباط الجزئي لعدد 12 حالة، مستخدماً اختبار أحادي الجانب:

rxy =.79

rAx =.81

rAx =.72

3- قام باحث بدراسة العوامل التي من الممكن أن تقود إلى التقليل من عدد أيام العمل والتي فقدت نتيجة للمرض، في مصنع معين. عشرة أشخاص تمت دراستهم. والبيانات التالية حول عدد الساعات التي يمارسون فيها التمارين الرياضية في الأسبوع، وكذلك أيام العمل التي غابوا فيها نتيجة لتعرضهم للمرض، يوضحها الجدول التالي:

عدد ساعات التمرين	عدد الأيام المفقودة	العمر بالسنوات
3	12	36
8	10	35
1	10	54
0	15	42
0	18	41
4	7	25
7	7	32
2	14	39
5	9	43
0	16	29
9	8	32
3	10	50

المطلوب:

- بين المتغير المعياري.
- ماذا تتوقع أن تكون الإشارة أمام المتغيرات المستقلة.
- ما هي معادلة الانحدار؟
- احسب معامل الانحدار المتعدد.
- هل تضمين العمر في النموذج يضيف شيئاً لقدرتنا على التنبؤ بعدد الساعات التي فقدت بسبب المرض.

4- بين كيف تم تطبيق تحليل الانحدار التدرجي في مجال البحث الاجتماعي، وما هي أسس هذا التطبيق؟

الهوامش والمصادر:

أولا: الهوامش:

1- Joseph F. Healey , The Essentials of Statistics: A Tool for Social Research , Wadsworth Cengage Learning , USA , 2010. pp. 362 - 364.

2- George DiEkhoff, Statictics for social and Behavioral Sciences: Univariat, Bivariate, Multivarite, wm. c. Brown Publishers , USA , 1992 , P. 261.

3- Ibid , P. 263.

4- George Argyrous , Statistics for Social and Health Research With Guide to SPSS , Sage Publications ,London , 2001 , P. 477.

5- Ibid , P. 479.

6- Ibid , P. 483.

7- Ibid , P. 484.

8- Ibid , P. 484.

9- Ibid , P. 485.

10- Ibid , P. 487.

11- Ibid , P. 487.

12- Joseph F. Healey , op. Cit. PP. 375 , 377 , 379.

ثانيا: المصادر:

1- George Argyrous , Statistics for Social and Health Research With Guide to SPSS , Sage Publications , London , 2001.

2- George Diekhoff, Statistics for Social and Behavioral Sciences, Univariate, Bivariate, Multivariate, wm. c. Brown Publishers, USA, 1992.

3- Joseph F. Healey, The Essentials of Statistitics: A Tool for Social Research, Wadsworth Cengate Learning, USA. 2010.

4- جولي بالانت (ت) خالد العامري، التحليل الإحصائي باستخدام برامج SPSS، ط 2، دار الفاروق للنشر والتوزيع، القاهرة، 2009.

5- عبد الله عامر الهمالي، التحديث الاجتماعي: معالمه ونماذج من تطبيقاته، منشورات جامعة قاريونس، 2008 م.

الجزء السادس

الإحصاءات اللابارامترية

- الفصل السابع عشر : اختبار مربع كاي (X2)لفروق الدلالة لعينة واحدة لتوزيع تكراري
- الفصل الثامن عشر : اختبار مربع كاي (X2) للاستقلال
- الفصل التاسع عشر : اختبار توزيع ذي الحدين لعينة واحدة
- الفصل العشرون : الأساليب الإحصائية للبيانات الترتيبية : اختبارات : مان وتني ، ولكوكسن ، كروسكال وليز وفريدمان

الفصل السابع عشر

اختبار مربع كاي (X^2) لفروق الدلالة لعينة واحدة لتوزيع تكراري

مقدمة:

إن كل الاختبارات التي تم تناولها في هذا الكتاب صممت في الأساس لاختبار الفروض حول معلمات محددة للمجتمع الإحصائي. فعلى سبيل المثال، لقد تم استخدام اختبارات t لتقييم الفروض حول u، ولاحقاً حول $u_1 - u_2$. إضافة إلى ذلك، أن هذه الاختبارات، على نحو نموذجي تقوم بافتراضات حول شكل توزيع المجتمع وحول معلمات أخرى للمجتمع. وبالرجوع بالذاكرة إلى تحليل الأنوفا ANONA، فإن توزيع المجتمع قد تم افتراض أنه موزع توزيعاً طبيعياً (أي اعتدالية التوزيع)، ومتجانس التباين، هي عملية مطلوبة، هي الأخرى، في هذا التحليل، ولما كانت هذه الاختبارات تهتم بالمعلمات وما تتطلبه هذه المعلمات من افتراضات، فقد أطلق عليها الاختبارات البارامترية Parametric Tests.

ومن الخصائص العامة الأخرى للاختبارات البارامترية أنها تتطلب وجود درجة عددية لكل حالة في العينة المدروسة. وتضاف الدرجات لبعضها البعض، ثم تُربَّع، ويُؤْخَذُ لها المتوسط. وبطريقة أخرى، تعالج باستخدام العمليات الحسابية الأساسية. أما

فيما يتعلق بموازين القياس، فإن الاختبارات البارامترية تتطلب بيانات مقاسة على المستويين ذي المسافات والنسبي.

في العادة يواجه الباحثون مواقف تجريبية قد لا تفي بمتطلبات الاختبارات البارامترية، أي عدم استيفائها لشروط الاختبارات المعملية. ففي مثل هذه المواقف، قد لا يكون من الملائم استخدام الاختبارات البارامترية. ومن هنا ينبغي على الباحث أن يضع في اعتباره أنه عندما لم يتم استيفاء شروط استخدام الاختبارات المعملية فإن إجراء الاختبار قد يقود إلى تفسيرات خاطئة للبيانات المستخدمة. من حسن الحظ أنه توجد تقنيات متعددة لاختبار الفرض كبدائل للاختبارات البارامترية. ويطلق على هذه البدائل الاختبارات غير البارامترية. في هذا الفصل سنتناول بعضاً من هذه الاختبارات غير البارامترية شائعة الاستخدام مثل مربع كاي (إحصاء يعرف بمربع كاي لحسن المطابقة ومربع كاي للاستقلالية) ويُسْتَخْدَمُ هذان الاختباران في تقييم الفروض حول التناسب، أو العلاقات الموجودة داخل المجتمعات الإحصائية. إضافة إلى اختبار ثنائي الحد كحالة خاصة لمربع كاي. وتجدر الإشارة إلى أن كلا الاختبارين المتعلقين بمربع كاي مثلهما مثل معظم الاختبارات غير البارامترية لا تتطلب افتراضات أو معلومات حول خصائص التوزيع الأساسي للمجتمع، وربما القيام ببعض الافتراضات (إن وجدت) حول توزيع المجتمع. ولذلك يطلق على الاختبارات غير البارامترية، في بعض الأحيان، اختبارات حرة التوزيع Distribution Free Tests.

إن أكثر الفروض وضوحاً بين الاختبارات البارامترية وغير البارامترية يكمن في نمط البيانات التي تستخدم في كل منهما.

إن كل الاختبارات التي تعاملنا معها حتى الآن، في متن هذا الكتاب، تتطلب درجات عددية Numerical Scores. أما على الجانب الآخر، فإن الاختبارات غير البارامترية، تتعامل مع بيانات.. هذه البيانات ينبغي أن تصنف على هيئة فئات مثل: عال، متوسط، منخفض، عصري، تقليدي... الخ، أي أنه يمكن ملاحظة أن هذه البيانات قد تكون مقاسة على المستوى الاسمي أو الترتيبي، وبالتالي لا تولّد لنا قيماً عدديةً يمكننا استخدامها في حساب المتوسطات الحسابية والتباين. وبدلاً من ذلك، فإن بيانات كثيرة من الاختبارات غير البارامترية، ببساطة بيانات تكرارية.

أحياناً. قد يلجأ الباحث إلى الاختيار بين استخدام الاختبار البارامتري والاختبار غير البارامتري. فاختيار الاختبار غير البارامتري، عادة يستلزم تحويل البيانات من درجات عددية إلى فئات غير عددية. فعلى سبيل المثال، يمكن للباحث أن يبدأ بدرجات عددية لقياس تقدير الذات Self Esteem، وتوليد ثلاث فئات تحتوي على:

درجة عالية لتقدير الذات، درجة متوسطة، ودرجة منخفضة.

وتجدر الإشارة: إلى أنه في معظم الأحوال، يفضل استخدام الاختبار البارامتري باعتباره أكثر ترجيحاً للكشف عن الفرق الحقيقي أو العلاقة الحقيقية. وعلى أية حال، تواجهنا مواقف يكون فيها تحويل الدرجات إلى فئات خياراً أفضل:

1- أحياناً، يكون من البساطة الحصول على مقاييس فئة. على سبيل المثال، من السهولة بمكان تصنيف الطلاب إلى طلاب ذي قدرات عالية، متوسطة أو منخفضة أكثر من الأداء على درجة عددية لقياس قدرة كل طالب على حدة.

2- إن الدرجات الأصلية قد لا تلبي بعض الافتراضات الأساسية لتؤكد إجراءات إحصائية محددة. على سبيل المثال، اختبار t واختبار أنوفا ANOVA يفترضان أن البيانات قد جاءت من توزيعات طبيعية Normal Distributions. كذلك، فإن المقاييس المستقلة للاختبارات هي الأخرى، تفترض أن المجتمعات المختلفة لديها نفس التباين (افتراضية التباين) The Homogeneity of Variance assumption. وعندما يشك الباحث في أن البيانات التي يتعامل معها، بيانات لا تفي بهذه الافتراضات، من هنا يكون من الأفضل تحويل الدرجات إلى فئات، وبالتالي، استخدام التقنيات الإحصائية اللامعلمية لتقييّم بياناته.

3- قد تكون الدرجات الأصلية في الواقع درجات لديها تباين عالٍ. ويعتبر هذا التباين المكون الأساسي للخطأ المعياري Standard Error لمقام معادلة إحصاء t، ومصطلح الخطأ Error Term في مقام معادلة نسبة F. F. Ratio. وعليه، فإن حجم التباين الكبير يمكن أن يقلل، وبشكل كبير، من احتمالية، أن هذه الاختبارات البارامترية سوف تكشف عن فروق دالة. وبتحويل الدرجات إلى فئات سيقلل جوهرياً من

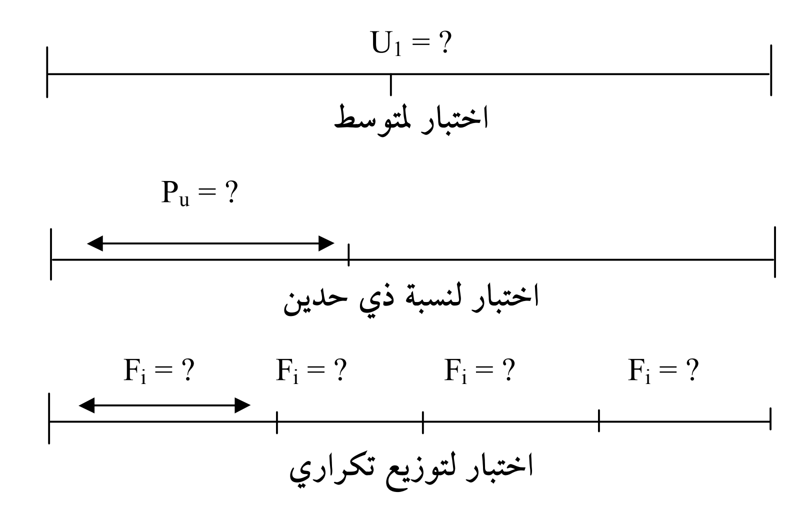
عملية التباين. فعلى سبيل المثال، أن كل الأفراد يتطابقون في الفئات الثلاث (عال، متوسط، منخفض) بغض النظر عن كيفما تكون الدرجات الأصلية[1].

ناقشنا في الفصول السابقة المواقف التي قمنا فيها باختبار فرض حول متوسط مجتمع أو نسبة لتوزيع ذي حدين. كما افترضنا أن متوسط مجتمع أو نسبة بقيمة محددة، وبعدئذ حددت احتمالية سحب العينة من هذا المجتمع بمتوسط أو نسبة تحصلنا عليها حقاً أثناء البحث. ونقوم بفعل ذلك من خلال حساب قيمة Z أو درجة t، ثم نقوم بإيجاد الاحتمالية المقابلة لهذه الدرجات من الجدول المعد لهذا الغرض (جدول توزيعات Z أو جدول توزيعات t). فإذا كان الفرق بين إحصاء العينة Sample Statistic، والقيمة المفترضة للمجتمع فرقاً كبيراً، فالاحتمالية المقابلة لتلك العينة التي تم سحبها من ذلك المجتمع سوف تكون منخفضة. وباختصار، فإن السؤال يختص بما إذا كان الفرق المشاهد بين إحصائية العينة والقيمة المفترضة للمجتمع كبيراً بشكل كافٍ.

اختبار مربع كاي لحسن المطابقة:

يمثل مربع كاي لحسن المطابقة اختبار لا معلمي لتوزيع الحالات عبر مدى واسع من القيم لمتغير واحد. وتستند فلسفة هذا الاختبار على حساب مقياس يعبر عن مدى الفرق بين أعداد المشاهدات والأعداد المتوقع مشاهدتها، فيما إذا كان النموذج الإحصائي صحيحاً. فإذا كان هذا المقياس صغيراً، كان النموذج مقبولاً والعكس بالعكس أي إذا كان هذا المقياس كبيراً فإنه بالتالي يتعذر قبول هذا النموذج الإحصائي[2].

إن طبيعة السؤال المطروح وفقاً لاختبار حسن المطابقة كمقابل لاختبارات أخرى التي تواجه الباحث يمكن توضيحها بالشكل التالي:

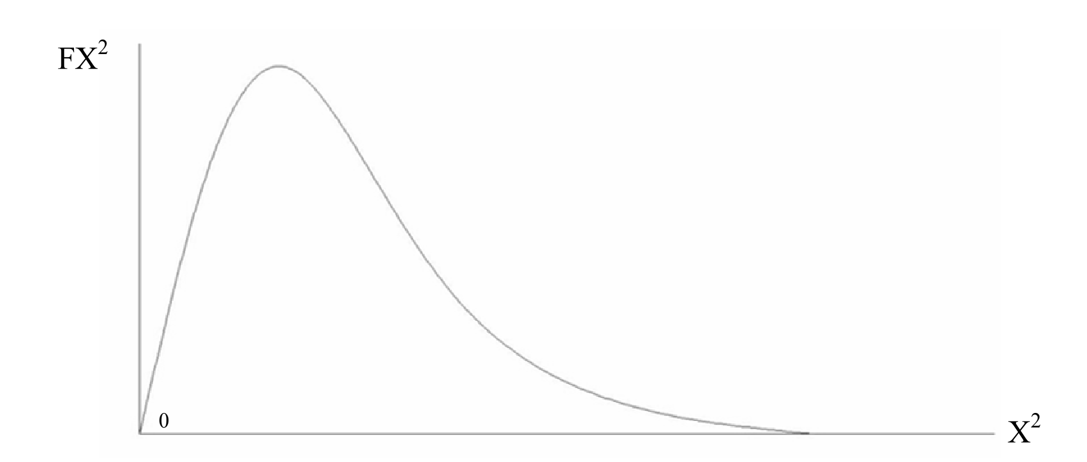


شكل (17ـ1) مقارنة بين اختبارات استدلالية

يحلل اختبار مربع كاي لحسن المطابقة التوزيع التكراري. وبما أن التوزيعات التكرارية يمكن بناؤها لبيانات اسمية وبيانات ترتيبية. كما هو الأمر للبيانات ذات المسافات المتساوية والنسبية. (كما بينا ذلك في الفصل الثالث). إذاً، اختبار مربع كاي لحسن المطابقة يمكن أن يستخدم لكل مستويات القياس. في هذا الفصل سنتعامل مع هذا الاختبار كما يطبق للبيانات الاسمية والبيانات الترتيبية حيث تنظم البيانات في فئات منفصلة. وبعد ذلك سوف نبين في سياق هذا الفصل تحليل التوزيع التكراري للبيانات المقاسة على المستويين ذي المسافات المتساوية والنسبي موضحين كيف أن هذا الاختبار يكون اختباراً مفيداً في تحليل التوزيع التكراري.

لقد أُطْلِقَ على هذا الاختبار اختبار مربع كاي لأن توزيع المعاينة الذي يتم استخدامه لتقييم الاحتمالية المتعلقة بالفرض الصفري على أنها صحيحة، هو توزيع مربع كاي (سوف نعود إلى هذه المسألة في الفصل المتعلق باختبار مربع كاي للاستقلال الأكثر استخداماً في مجال البحوث الاجتماعية).

إن الصورة العامة لتوزيع مربع كاي تتضح من خلال الشكل التالي:



شكل (17 ـ 2) توزيع مربع كاي (X^2)

ويتخذ بناء توزيع مربع كاي نفس الأساس كما في توزيعات المعاينة التي تمت مناقشتها. ونعني بتوزيعات المعاينة احتمالية توزيع إحصاء العينة التي نتحصل عليها من عدد لا متناهي من العينات ذات الأحجام المتساوية المسحوبة من مجتمع بخصائص محددة.

ولتوضيح اختبار حسن المطابقة نحاول الإجابة على السؤال التالي:

هل يتأثر معدل الجريمة بفصول السنة؟ عند طرح هذا السؤال فإن الباحث لا يرغب بشكل واضح في معرفة معدل الجريمة بقدر ما يتركز اهتمامه على توزيعات معدلات الجريمة عبر فصول السنة.

الإجراءات:

بادئ ذي بدء ينبغي على الباحث صياغة فرض حول توزيع المجتمع: يفترض أنه لا توجد علاقة بين معدلات الجريمة وفصول السنة. ومن خلال هذا الفرض يتوقع الباحث أن عدد الجرائم التي ارتكبت في أي عام تُوَزَّع بشكل متساوٍ عبر الفصول الأربعة.

$$Fe = \frac{\text{العدد الإجمالي للجرائم (N)}}{4}$$

حيث إن Fe تعني التوزيعات التكرارية في كل فئة.

من ناحية ثانية، يمكننا القول، بأن معدل الجريمة في أي سنة قد يتأثر بأحداث عشوائية تسبب فرقاً طفيفاً في التوزيع عن تلك النتيجة المتوقعة. بمعنى آخر، أنه ليست كل عينة من العينات ستتطابق مع هذا التوقع لعدد متساوٍ من الجرائم المرتكبة في كل فصل من فصول السنة. ويمكننا توضيح الفرق بين القيمة المتوقعة والقيمة المشاهدة بحساب إحصاء عينة كاي تربيع:

$$x^2 = \sum \frac{(Fo - Fe)^2}{Fe} \quad \text{العينة}$$

حيث إن: Fe : التوزيعات المتوقعة في كل فئة.

F0 : التوزيعات المشاهدة في كل فئة.

تجدر الإشارة إلى القول بأنه إذا كانت نتيجة العينة تطابق تماماً النتيجة المتوقعة، فإن قيمة مربع كاي للعينة تطابق تماماً النتيجة المتوقعة، فإن قيمة مربع كاي للعينة ستكون صفراً(0).

إن السؤال الذي يمكن طرحه هو: ما هو الموقف عندما يكون التوزيع المشاهد ليس على نحو دقيق مساوياً للتوزيع المتوقع؟

بالنظر إلى معادلة X^2 يمكننا مشاهدة أن أي فرق سيولد قيمة إيجابية لعينة مربع كاي. وهذا سببه أن أي فرق قد تم تربيعه. وبذلك يمكننا التخلص من القيم السالبة. كذلك يمكننا ملاحظة أنه كلما كان الفرق كبيراً بين التوزيعات المشاهدة والتوزيعات المتوقعة (البواقي Residuals)، كانت القيمة عالية (موجبة) لعينة مربع كاي. إن السؤال إذاً يصبح حول ما هي النقطة الأساسية وهي هل القيمة المتعلقة بعينة مربع كاي تصبح كبيرة جداً إلى درجة أن هذه القيمة تقترح أن العينة لم يتم اختيارها من مجتمع تنتشر فيه معدلات الجريمة بشكل متسق عبر فصول السنة؟

يمكننا اختيار القيمة الحرجة لـِ X^2 بنفس الطريقة المتبعة في الاختبارات التي تناولناها في الفصول السابقة. وكذلك يمكننا اختيار مستوى الفا (α) مثل 0.05، كما يمكننا

استخدام جدول توزيع قيم مربع كاي الحرجة (انظر الملحق) لإيجاد القيمة الحرجة لمربع كاي بمستوى معين من الفا آخذين في الاعتبار عدد درجات الحرية (df) لأي توزيع معطى، وللحصول على درجة الحرية تستخدم المعادلة التالية:

$$df = k - 1$$

حيث إن: K عدد الفئات

وعليه، إذا كان لدينا متغير يتألف من أربع فئات كما هو الحال في المثال الذي نتعامل معه (معدلات الجريمة حسب فصول السنة)، فإن درجة الحرية ستكون كالتالي:

$$df = 4 - 1 = 3$$

ومن خلال مراجعة الجدول المتعلق بتوزيعات القيم الحرجة لمربع كاي بمستوى الفا 0.05 بدرجة حرية 3، فإن قيمة مربع كاي الحرجة تساوي

$$x^2 = 7.815(\alpha = 0.05, df = 3)7.815$$

الآن، يمكننا حساب قيمة العينة لمربع كاي (X^2) لنرى ما إذا كانت قيمة العينة تقع داخل المنطقة الحرجة (منطقة الرفض)، على سبيل المثال، إذا لاحظنا فعلياً توزيع (الافتراضي) الجريمة كما هو مبين في الجدول (1). فالسؤال هو: هل يمكننا أن نخلص إلى القول بأن الجريمة متأثرة بفصول السنة؟

جدول (17 ـ 1) توزيع الجريمة حسب فصول السنة

	الصيف	الربيع	الشتاء	الخريف	المجموع
المشاهد	300	270	200	250	1020
المتوقع	255	255	255	255	1020
البواقي	45	15	−55	−5	

المصدر: .George Argyrous, Statistics for Social and Healh Research, op. Cit , P.325

للحصول على القيم المتوقعة يمكننا استخدام المعادلة التالية:

تقسيم المجموع الكلي على عدد الفصول. = $\frac{1020}{4} = 255$

إن الصف الذي يطلق عليه البواقي في الجدول يعني الفرق بين القيم المشاهدة والقيم المتوقعة.

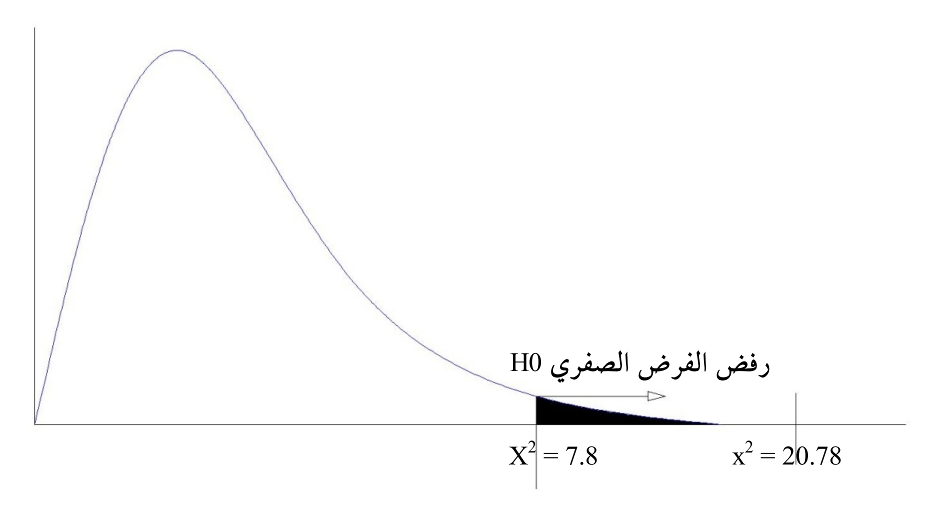
الآن يمكننا أن نستعيض عن نتائج العينة إلى المعادلة المتعلقة بمربع كاي:

العينة $X^2 = \sum \frac{(Fo - Fe)^2}{Fe}$

$$= \frac{(300-255)^2}{255} + \frac{(270-255)^2}{255} + \frac{(200-255)^2}{255} + \frac{(250-255)^2}{255}$$
$$= 20.78$$

والشكل التالي يوضح درجات العينة والقيم الحرجة X^2 لمساعدتنا في اتخاذ القرار.

ومن خلال هذا الشكل، تبين لنا أن قيمة مربع كاي للعينة بشكل واضح تقع داخل المنطقة الحرجة، الأمر الذي يقودنا إلى رفض الفرض الصفري للتوزيع المتساوي للجريمة عبر فصول السنة [3].



شكل (17 ـ 3) توزيع (X^2) ودرجة الحرية (df = 3)

إجراء اختبار مربع كاي لحسن المطابقة باستخدام Spss :

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Chi - Square ⇐ Non Parametric tests ⇐ Analyze

2- انقر على Crime by Season في القائمة.

3- انقر على ◄ يقوم بلصق Crime by Season في Test Variable List.

4- انقر على ok.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

Chi - Square Test

Frequencies

	Observed (N)	Expected (N)	Residual
Summer	300	255.0	45.0
Spring	270	255.0	15.0
Autumn	200	255.0	-55.0
Winter	250	255.0	-5
Total	1020		

Test Statistics

	Crime by Season
Chi - Square[a]	20.784
df	3
Asymp. sig	000

a. o cells (o %) have expected frequencies less than 5
The Minumum expected cell frequencies is 255.0

المصدر: .George Argyrous, Statistics for Socialand Healh Research, op. Cit , P.328

شكل (17 ـ 4) مخرجات Spss لاختبار مربع كاي

تفسير مخرجات SPSS لمربع كاي (X^2) :

يحتوي المربع المُعَنْوَن: الجريمة وفقاً لفصول السنة Crime by Season على الإحصاء الوصفي الذي يلخص العينة المدروسة. وفي هذه الحالة فإن توزيع الحالات عبر الفصول الأربعة من فصول السنة قد يوفرها عمود التوزيعات المتوقعة التي تولد لنا العدد المتساوي لكل الحالات المتوقعة في كل فصل من فصول السنة.

إن القيم المتوقعة لعمود (N) قد طرحت من عمود (N) القيم المشاهدة تعطينا الفرق بين القيم المشاهدة والقيم المتوقعة.

أما المربع الثاني المُعَنْوَنُ بـِ إحصاء الاختبار لمربع كاي Chi - Square Test Statistics، نجد أن قيمة مربع كاي تساوي 20.784 بدرجة حرية 3 df(3) بدلالة تصل إلى 1 في كل 10.000 عينة إذا ما وزعت الجريمة بشكل متساوٍ عبر فصول السنة. إن مثل هذه الاحتمالية المنخفضة تقودنا إلى رفض الفرض الصفري: معدلات الجريمة يبدو أنها مرتبطة بفصول السنة[4].

اختبار مربع كاي لحسن المطابقة للحالة السوية (الاستواء Normality) :

من خلال المثال الذي تعاملنا معه في اختبار مربع كاي لحسن المطابقة للبيانات المقاسة على المستوى الاسمي التي تقع ضمن فئات منفصلة. وفي هذا السياق يمكننا القول بأن أي اختبار يمكن تطبيقه على بيانات اسمية وبيانات ترتيبية مع ذلك، يمكن تطبيقه على مستويات أعلى من المقياس كالبيانات ذات المسافات المتساوية والنسبية. ففي حالة المقياس ذي المسافات والنسبي فإننا ننظر إلى التوزيع التكراري للحالات عبر مدى من القيم أو عبر فئات متساوية تماماً بنفس الطريقة التي ينظر فيها إلى التوزيع عبر فئات منفصلة.

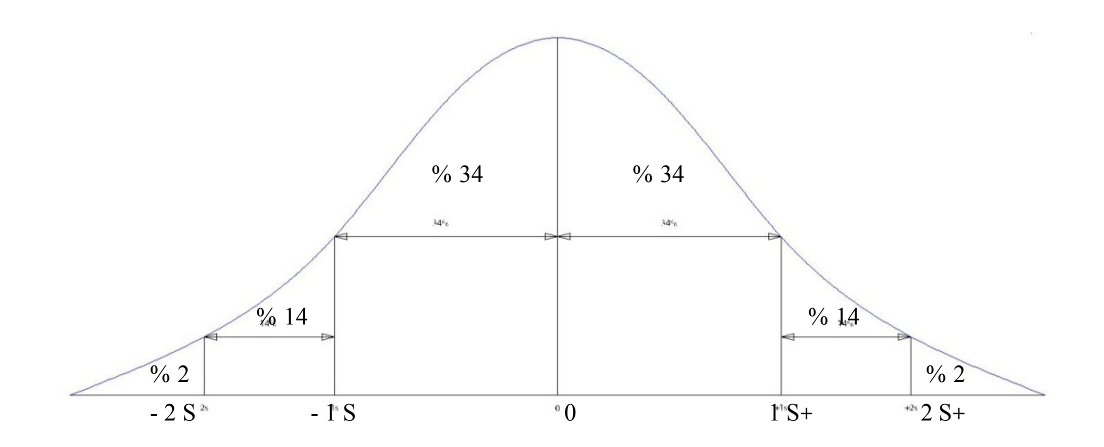
إن هذا هو المنطق الذي يجعل من اختبار حسن المطابقة مقياساً مفيداً في تقييم ما إذا كانت البيانات المقاسة على المستويين ذي المسافات والنسبي جاءت من مجتمع طبيعي. من هذه الناحية، فإن هذا الاختبار هو اختبار لا معلمي Non - Parametric Test يمكن أن يكون مفيداً كاختبار تمهيدي وتكملة للاختبارات المعلمية Parametric Tests التي تتطلب الافتراض أن العينة سحبت من مجتمع طبيعي [5].

مثال: نفترض أن لدى الباحث عينة ويرغب في تقييم ما إذا كان قد تم سحبها من مجتمع طبيعي. وكما سبقت الإشارة، فقد عرف التوزيع الطبيعي بالتوزيعات التكرارية كما هو مبين في الجدول رقم (2) وشكل رقم (5).

جدول (17 ـ 2) توزيع المنحنى الطبيعي

مدى القيم Range of Value	نسبة الحالات Percentage of Cases
أبعد من 2 انحراف معياري تحت المتوسط	% 2
بين 1 و 2 انحرافات معيارية تحت المتوسط	% 14
داخل 1 انحراف معياري تحت المتوسط	% 34
داخل 1 انحراف معياري فوق المتوسط	% 34
بين 1 و 2 انحرافات معيارية فوق المتوسط	% 14
أبعد من 2 انحرافات معيارية فوق المتوسط	% 2

المصدر: George Argyrous, Statistics for Social and Healh Research, op. Cit , P.331.



شكل (17 ـ 5) المساحات تحت المنحنى الطبيعي

إنه بإمكاننا استخدام القيم النسبية كما هو مبين في الجدول رقم (17 - 2) لحساب القيم المتوقعة باستخدام المعادلة المتعلقة بمربع كاي. لاحظ مرة أخرى أن هذا المثال لا يشبه المثال السابق حول معدل الجريمة حسب فصول السنة، حيث إننا لا نفترض هنا أن الحالات موزعة توزيعاً متساوياً عبر الفئات، وإنما بدلاً من ذلك فالتوزيعات المتوقعة قد استندت على خصائص المنحنى الطبيعي.

مثال(6):

نفترض أن لدى الباحث عينة مؤلفة من 110 أشخاص بمتوسط عمري يصل إلى 45 سنة، وبانحراف معياري 10 سنوات، إذا كانت هذه العينة موزعة توزيعاً طبيعياً، فالباحث يتوقع أن يجد عدد الأشخاص داخل المدى المبين الذي يبينه الجدول رقم (17 - 3).

جدول (17 - 3) التوزيع المتوقع للعينة

مدى القيم	نسبة الحالات	عدد الحالات
25 سنة أو أقل (أبعد من 2 انحرافات معيارية تحت المتوسط)	% 2	$2.2 = 110 \times \frac{2}{100}$
26 - 35 سنة (بين 1 و 2 انحرافات معيارية تحت المتوسط)	% 14	$15.4 = 110 \times \frac{14}{100}$
36 - 45 سنة (داخل 1 انحراف معياري تحت المتوسط)	% 34	$37.4 = 110 \times \frac{34}{100}$
46 - 55 سنة (داخل 1 انحراف معياري فوق المتوسط)	% 34	$37.4 = 110 \times \frac{34}{100}$
56 - 65 سنة (بين 1 و 2 انحراف معيار فوق المتوسط)	% 14	$15.4 = 110 \times \frac{14}{100}$
66 سنة أو أكبر (أبعد من 2 انحراف معياري فوق المتوسط)	% 2	$2.2 = 110 \times \frac{2}{100}$

من ناحية ثانية، يمكننا فعلياً الحصول على توزيعات عينة كما هو مبين في الجدول التالي:

جدول (17 ـ 4) التوزيعات المشاهدة للعينة

مدى القيم	عدد الحالات
25 سنة أو أقل	5
26 - 35 سنة	17
36 - 45 سنة	33
46 - 55 سنة	33
56 - 65 سنة	17
66 سنة أو أكبر	5

من خلال المقارنة بين القيم المشاهدة والقيم المتوقعة يتجلى لنا بوضوح الفرق بين القيم المشاهدة والقيم المتوقعة: والسؤال الذي يُطْرَح في هذا السياق هو: هل هذا الفرق يقودنا إلى رفض الفرض الذي مفاده أن المجتمع موزع توزيعاً طبيعياً؟

للإجابة على هذا السؤال، يتطلب منا حساب قيمة مربع كاي(X^2):

$$x^2 = \sum\frac{(Fo-Fe)^2}{Fe} \quad \text{العينة}$$

$$= \frac{(5-2.2)^2}{2.2} + \frac{(17-15.4)^2}{15.4} + \frac{(33-37.4)^2}{37.4} + \frac{(33-37.4)^2}{37.4} + \frac{(17-15.4)^2}{15.4} + \frac{(5-2.2)^2}{2.2}$$

$$= 8.5$$

بعد أن أجرينا إحصاء اختبار مربع كاي، ينبغي علينا مقارنة درجة العينة مع القيمة الحرجة لمربع كاي التي تستند في الأساس على درجة الحرية (df). في هذا المثال، لدينا

ست فئات، وبالتالي فإن درجة الحرية تساوي 5 = 1 – 6، على مستوى ألفا (α) 05.، والقيمة الحرجة لمربع كاي هي 11.070 (انظر ملحق توزيع القيم الحرجة لمربع كاي).

ولما كانت القيمة المحسوبة لمربع كاي أصغر من القيمة الحرجة، عليه، لا يمكننا رفض الفرض الصفري الذي مفاده أن العينة جاءت من مجتمع موزع توزيعاً طبيعياً.

خلاصة:

لقد تم مناقشة اختبار X^2 لحسن المطابقة. وبالرغم من أن اختبار X^2 مرتبط بحسابات مختلفة قليلاً، إلا أنه يشبه إلى حد كبير الاختبار "ذا الحدين" الذي سنتناوله لاحقاً. حيث أن اختبار Z لتوزيع ذي الحدين للنسب يطبق فقط لتوزيعات تكرارية تم تنظيمها في توزيع ذي حدين.

إن اختبار مربع كاي يكون اختباراً أكثر شمولاً عند تطبيقه على التوزيعات التكرارية لأي عدد من الفئات (وبالتالي فإن اختبار Z لتوزيع ذي حدين للنسب يمكن اعتباره حالة خاصة لاختبار X^2.

إن هذا الوضع يعطي اختبار X^2 تطبيقات واسعة إلى عينتين أو أكثر كما سنرى في الفصل المتعلق باختبار مربع كاي للاستقلال.

أسئلة للمراجعة:

1- ما هي درجات الحرية والقيم الحرجة لـِ X^2 لكل من:

الفا 0.10 = α و الفا 0.05 = α لاختبار حسن المطابقة لمتغير يحتوي على:

أ- ثلاث فئات ب- خمس فئات ج- ثمان فئات.

2- احسب اختبار حسن المطابقة للبيانات التالية لاختبار الفرض الذي مفاده: أن العينة التي تم سحبها من مجتمع ذي توزيع متماثل لحالات عبر الفئات:

أ-

القيم	عدد الحالات
1	45
2	40
3	55
4	54
5	38

ب-

القيمة	عدد الحالات
1	120
2	111
3	119
4	125
5	120
6	127
7	118

3- تم مسح 90 شخصاً لمعرفة كمية الوقت الذي يتم استغلاله في القراء كل يوم، وتم قياس كمية الوقت. يرغب الباحث في اختبار الفرض الذي مفاده أن العينة التي تم سحبها من مجتمع موزع توزيعاً طبيعياً. علماً بأن متوسط القراءة للعينة المسحوبة

يصل إلى 45 دقيقة، بانحراف معياري 15 دقيقة. وأن توزيع هذه القيم المشاهدة تظهر بالشكل التالي:

مدى القيم	عدد الحالات
أقل من 16 دقيقة	3
16 – 30 دقيقة	15
31 – 45 دقيقة	34
46 – 60 دقيقة	31
61 – 75 دقيقة	5
أكثر من 75 دقيقة	2

- مستخدماً ألفا 0.05 = α لاختبار فرضية الاستواء (الحالة السوية) للمجتمع، أجر اختبار حسن المطابقة؟

4- تم مقارنة خمس مدارس فيما يتعلق بنسبة الطلاب الذين يتابعون تعليمهم الجامعي. عينة من 50 طالباً قد تخرجوا من كل مدرسة تم سحبهم. وعدد أولئك الذين دخلوا الجامعة من كل مدرسة هم:

المدرسة	العدد الذي دخل الجامعة
مدرسة 1	22
مدرسة 2	25
مدرسة 3	26
مدرسة 4	28
مدرسة 5	33

المطلوب:

أ- حساب القيم المتوقعة، وبعد ذلك حساب قيمة X^2 لاختبار حسن المطابقة.

ب- ما هي النتائج التي توصلت إليها حول توقع دخول الجامعة من كل مدرسة من هذه المدارس؟

الهوامش والمصادر:

أولا: الهوامش:

1. Frederick J. Gravetter and Larry B. Wallnau , Statistics for The Behavioral Sciencies , Wadsworth Cengage Learning , USA , 2010 , PP. 606 - 607.
2. George Argyrous, Statistics for Social & Health Research, with Aguide to Spss , SAGE Publications, London, 2001 , P. 323.
3. Ibid., P. 327.
4. Ibid., P. 331.
5. Ibid., P. 331.
6. Ibid., PP. 332 - 334.

ثانيا: المصادر:

1- Frederick J. Gravetter and Larry B. Wallnau , Statistics for The Behavioral Sciencies , Wadsworth Cengage Learning , USA , 2010.

2- George Argyrous, Statistics for Social & Health Research, with aguide to Spss , SAGE Publications, London, 2001.

3- Hugh Coolicam , Hodder Arnold , Research Methods and Statistics in Psychology , 4th ed , Hodder Stoughton Educational , London , 2004.

الفصل الثامن عشر
اختبار مربع كاي (X^2) للاستقلال

مقدمة:

نتناول في هذا الفصل أحد التقنيات الإحصائية التي يمكن من خلالها إجراء اختبار فرضية لبيانات مقولية تم ترتيبها في جداول متقاطعة. ويطلق على هذه التقنية الإحصائية اختبار مربع كاي (X^2) للاستقلال. ويشبه اختبار مربع كاي للاستقلال اختبار عينة واحدة التي تم تناولها في موضع آخر من هذا الكتاب.

اختبار مربع كاي واختبارات الدلالة الأخرى:

في الفصول السابقة لهذا الكتاب تم التركيز على اختبار الاستدلال الذي تم تحديده من خلال معيارين أساسيين أولهما: الإحصاء الوصفي المستخدم في وصف البيانات الخام لعينة. وثانيهما: عدد العينات التي تم وصفها من خلال الاستدلال الذي تم القيام به.

1- الإحصاء الوصفي المستخدم في وصف البيانات الخام:

إن هذا العامل في حد ذاته دالة لقضيتين أساسيتين: أولهما: متعلقة بالسؤال البحثي المطروح الذي يتعين على الباحث الإجابة عليه. إن السؤال البحثي تقريباً وبشكل ثابت يوجه اهتمامات الباحث نحو خصائص محددة لتوزيع متعلق بالمتغير الخاضع للاستقصاء. على سبيل المثال، أن الباحث في شئون الصحة العامة، قد يَنْصَبُّ اهتمامُه حول سؤال مفاده ما إذا كان مجتمع ما يكون فيه المتوسط العمري يميل نحو الشباب أو الكبار. إذاً مشكلة البحث هذه تقود الباحث إلى استخدام إحصاءات النزعة المركزية لهذا المتغير. وكذلك الأمر بالنسبة للباحث في العلوم السياسية الذي قد ينصب اهتمامه أيضاً على معرفة التوزيع العمري لهذا المجتمع، إلا أنه في هذه الحالة فإن التركيز يمكن أن ينصب على العينة العمرية التي يحق لها المشاركة في العملية السياسية. من هنا نجد أن الباحث سيقوم بتنظيم بياناته من خلال توزيع ذي حدين وحساب نسبة العينة الأعلى والأدنى للفئة العمرية. إن كلا الباحثين يرغبان في التعامل مع نفس المجتمع، وإن كليهما، لديه نفس البيانات الخام. إلا أن السؤال البحثي المطروح هو الذي يحدد ما إذا كان الباحثان يرغبان في تبني توزيعات النزعة المركزية أو نسب الحالات الأعلى والأدنى من نقطة محددة على المقياس. وثانيهما يتعلق بالقضية التي تحدد الإحصاء الوصفي الذي يستخدم لتلخيص البيانات وهي مستوى القياس للمتغير. إن مستوى القياس كثيراً ما يقيد الطريقة التي يمكن أن نصف بها خصائص التوزيع الذي يرغب الباحث فيه. على سبيل المثال، نفترض أننا نرغب في معرفة النزعة المركزية للتوزيع العمري لعينة ما، وسواء قمنا بجمع بيانات مقاسة على المستوى ذي المسافات والنسبي، أو مقاسة على المستوى الترتيبي فهي التي تحدد مقاييس النزعة المركزية (المنوال، الوسيط، أو المتوسط) فنحن في الواقع نستخدم هذه الإحصاءات الوصفية لتلخيص البيانات الخام. فإذا تم قياس العمر بالسنوات (مقياس ذو المسافات المتساوية والنسبي)، فإنه في هذه الحالة يعتبر الإحصاء الوصفي المستخدم لتلخيص البيانات هو المتوسط الحسابي. أما إذا تم تصنيف العمر حسب الفئات في مدى يتراوح من: صغير جداً إلى كبير جداً (مقياس ترتيبي) إذاً في هذه الحالة فإن الإحصاء الوصفي المستخدم هو الوسيط. إن هذين العاملين المتمثلين في

السؤال البحثي ومستوى القياس معاً يحددان الإحصاء الوصفي الفعلي الذي على ضوئه يتم إجراء الاختبارات الاستدلالية.

عدد العينات التي ينبغي مقارنتها:

كما رأينا في الفصل المتعلق باختبار t لمتوسط حسابي لعينة واحدة، فإنه عندما يتم جمع البيانات من عينة واحدة فقط، حينئذ يكون لدينا مدى معين لاختبارات الاستدلال للاختيار من بينها: أنه يمكن القول بأن مدى هذا الاختيار يختلف عند التعامل مع عينتين، ولذلك ينبغي علينا في هذه الحالة أن نقوم باستدلال حول كل واحد من هذين المجتمعين اللذين سُحِبَتْ منهما هاتان العينتان. وبالتماثل، إذا كانت لدينا أكثر من عينتين، فإننا في هذه الحالة نواجه مدى آخر من الاختبارات للاختيار من بينها. على سبيل المثال، عند مقارنة متوسطات باختبار t لمتوسطات العينة يكون ملائماً لعينة أو عينتين. في حين أن تحليل ANOVA يستخدم لأكثر من عينتين وهكذا. واضعين هذه القضايا بعين الاعتبار، فإنه بإمكاننا النظر إلى الأحوال التي يكون فيها استخدام مربع كاي ملائماً.

إن الإحصاء الوصفي الذي على ضوئه يستخدم مربع كاي للاستقلال هو التوزيع التكراري المتضمن في جدول ثنائي bivariate table. في الفصل السابع من هذا الكتاب قد بيّنا عملية بناء واستخدام الجداول الثنائية أي جداول التقاطع. وتعد جداول التقاطع جداول ملائمة في تلخيص وعرض البيانات المقولية عند الرغبة في التوزيع الكلي للحالات عبر مجموعة المدى الكلي للفئات مفضلاً ذلك أو بدلاً من مقاييس النزعة المركزية. وتعتبر جداول التوافق ذات أهمية كبيرة وذلك لشيوع استخدامها في وصف البيانات المقاسة على المستوى الاسمي والترتيبي.

وتجدر الإشارة إلى أن البيانات المقاسة على المستويين ذي المسافات والنسبي يمكن اختزالها إلى فئات منفصلة مثل اختزال الدخل إلى فئات: ذوي الدخل المنخفض والمتوسط، والعالي. كما تعتبر جداول التقاطع أداة مهمة لوصف البيانات ذات المسافات المتساوية والنسبية كما هو الحال في أهمية البيانات الاسمية والترتيبية.

إن اختبار مربع كاي يعتبر في الأساس الشيء نفسه بغض النظر عما إذا كان لدينا

عينة واحدة، أو عينتان أو أكثر من ذلك. وكما بينا في الفصل المتعلق باختبار مربع كاي لعينة واحدة لتوزيع تكراري. إن مربع كاي، ليس كالاختبارات الأخرى، فهو اختبار يمكن توسيعه ليشمل عينتين أو أكثر من عينتين من الحالات دون تعديلات كبيرة: فالباحث يمكنه اتباع الإجراء نفسه، والمعادلة نفسها، بغض النظر عن عدد العينات الخاضعة للمقارنة (لاحظ أنه في حالة العينة الواحدة، فقد أطلق على مربع كاي اختبار حسن المطابقة A goodness-of-fit-test في حين يطلق على مربع كاي للاستقلال A test for Ivdependence، عند التعامل مع عينتين أو أكثر).

الاستقلال الإحصائي Statistical Independence:

يمدنا مفهوم الاستقلال الإحصائي بطريقة أخرى تمكننا من معرفة الفرق بين الفرض البحثي والفرض الصفري عندما يكون لدينا متغيران مستقلان إحصائياً فإن أي تغير في واحد من المتغيرين ليس له علاقة بالتغير في المتغير الآخر لأنهما يختلفان بشكل مستقل عن بعضهما البعض. أما على الجانب الآخر، إذا كان هذان المتغيران يعتمدان على بعضهما البعض فإن أي تغير في أحدهما يرتبط بالتغير في المتغير الآخر.

مثال:

دعنا نعيد العمل من خلال المثال الذي أوردناه في الفصل السابق لبناء جدول تقاطعي بين متغيري (الدخل ومكان الإقامة).

جدول (18ـ 1) العلاقة بين الدخل ومكان الإقامة

مكان الإقامة	الدخل		
	منخفض	عال	المجموع
ريفي	205 (%55)	118 (%34)	323 (%45)
حضري	167 (%45)	230 (%66)	397 (%55)
المجموع	372 (%100)	348 (%100)	720 (%100)

المصدر: George Argyrous, Statistics for Social & Health Research, op.cit, p.395.

بالنظر إلى التوزيعات النسبية لكل خلية من خلايا الجدول يمكننا تقييم ما إذا كان في العينة أن المتغيرين مستقلان، أو ما إذا كان في الحقيقة أنه يُوجد نوع من العلاقة. ومن خلايا بيانات هذا الجدول نلاحظ وجود بعض العلاقة (أو نمط الاعتماد Pattern of Independence) بين هذين المتغيرين. ويبدو من خلال الأرقام المبينة في الجدول أعلاه، أن الأفراد ذوي الدخول المنخفضة يتجهون إلى العيش في المناطق الريفية. في حين أن الأفراد ذوي الدخول المرتفعة يميلون إلى العيش في المناطق الحضرية.

إضافة إلى جدول التقاطع. فقد تم حساب مقاييس التطابق الملائمة لتكميم هذه العلاقة المشاهدة بين هذين المتغيرين. ولما كان على الأقل أحد هذين المتغيرين تم قياسه على المستوى الاسمي، فإن المقياس المناسب هو لامبيدا Lambda وأن قيمة لامبيدا لهذا الجدول تساوي 0.12 وتشير هذه الدرجة إلى علاقة ضعيفة لهذه البيانات.

ومن ناحية ثانية، فإن النتيجة التي توصلنا إليها قد اعتمدت على بيانات عينة، وعليه ينبغي علينا أن نكون حذرين من أن هذه النتيجة ربما قد جاءت نتيجة للتباين العشوائي عند عملية المعاينة من المجتمعات التي لا توجد فيها أي علاقة بين الدخل ومكان الإقامة. وأن اختبار مربع كاي كفيل بتقييم هذه الاحتمالية.

حساب وتفسير إحصاء الاختبار لجداول مربع كاي للاستقلال :

يمتلك مربع كاي عنصرين أساسيين هما إحصاء الاختبار الذي يطلق عليه القيمة المتحصل عليها لمربع كاي. والعنصر الثاني القيمة الحرجة للإحصاء التي تمثل القيمة لإحصاء الاختبار الذي ينبغي أن يمتد لرفض الفرض الصفري.

إن إحصاء الاختبار لمربع كاي ليس مقياساً للتطابق لأن قيمة مربع كاي تكون متأثرة بالدرجة التي يكون فيها المتغيران معتمدين على بعضهما البعض. إن قيمة مربع كاي سوف تكون أكبر إذا كان المتغير التابع يعتمد على المتغير المستقل. في حين تكون قيمة مربع كاي أصغر إذا كان المتغير التابع يختلف بشكل مستقل بدون النظر إلى المتغير المستقل[1].

إن أول خطوة لإجراء مربع كاي للاستقلال، كما هو الحال، في كل مقاييس الاستدلال هي البدء بصياغة الفرض الصفري والفرض البديل، ففي المثال الذي أمامنا، فإن صياغة الفرض الصفري تتخذ الشكل التالي:

Ho: الدخل ومكان الإقامة متغيران مستقلان عن بعضهما البعض.

Hi : الدخل ومكان الإقامة ليسا متغيرَيْن مستقلين عن بعضهما البعض.

إذا رفضنا الفرض الصفري الذي مفاده أن المتغيرين مستقلان، فإننا في هذه الحالة نستنتج أن المتغيرين ليسا مستقلين في المجتمع المدروس. وعلى العكس من ذلك، إذا عجزنا عن رفض الفرض الصفري، فإننا نجادل في أن المتغيرين حقاً هما مستقلان بالرغم من أن التبعية قد تم ملاحظتها في العينات.

إذا نظرنا إلى مثالنا الحقيقي، فإننا نكون قد حددنا أن المتغيرين ليسا مستقلين في العينة - أي لا تبدو أن هناك بعض العلاقة - ولكن هل باستطاعتنا استخلاص هذا الاستدلال حول المجتمعات التي سحبت منها هذه العينات؟.

ولكي نرى كيف يساعدنا اختبار مربع كاي في تقييم ما إذا كان هذان المتغيران حقاً مستقلين عن بعضهما البعض بالرغم من وجود تبعية في العينات. دعنا نبدأ بالنظر إلى البيانات الخام والتوزيعات النسبية في الجدول التالي:

جدول (18 ـ 2) مكان الإقامة: كل المبحوثين

مكان الإقامة	المجموع	النسبة المئوية
ريفي	323	% 45
حضري	397	% 55
المجموع	720	% 100

إن مجاميع الصف والنسب المئوية هي النقاط المرجعية التي من خلالها يتم إجراء اختبار مربع كاي. إن الحجة تكون أنه إذا كان 45% من مجموع المبحوثين يعيشون في

مناطق ريفية، حينها نتوقع أن 45% من كل مجموعة (دخل منخفض، ودخل عالٍ) هم أيضاً يعيشون في مناطق ريفية إذا كان المتغيران مستقلين.

الجدول التالي يوضح هذه التوزيعات النسبية المتوقعة.

جدول (18 ـ 3) التوزيعات النسبية المتوقعة للخلية

مكان الإقامة	الدخل		
	منخفض	عال	المجموع
ريفي	% 45	% 45	% 45
حضري	% 55	% 55	% 55
المجموع	% 100	% 100	% 100

من ناحية ثانية، حتى ولو أن الفرضية الصفرية للاستقلال صحيحة، فإنه ليس في كل الأحوال أن نتوقع دائماً عينات عشوائية لأولئك الذين يكسبون دخولاً منخفضة ودخولاً عالية لتعكس ذلك. فعلى سبيل المثال، ربما أحياناً قد نسحب عينات من مجموعات ذوي دخول منخفضة وعالية ونحصل على واحدة من النتائج الثلاث التالية كما هو موضح في الجدول التالي:

جدول (18 – 4 أ) عينة 1

مكان الإقامة	الدخل		
	منخفض	عال	المجموع
ريفي	% 47	% 44	% 45
حضري	% 53	% 56	% 55
المجموع	% 100	% 100	% 100

جدول (18 – 4 ب) عينة 2

	الدخل		
مكان الإقامة	منخفض	عال	المجموع
ريفي	% 52	% 40	% 45
حضري	% 48	% 60	% 55
المجموع	% 100	% 100	% 100

جدول (18 – 4 ج) عينة 3

	الدخل		
مكان الإقامة	منخفض	عال	المجموع
ريفي	% 65	% 30	% 45
حضري	% 35	% 70	% 55
المجموع	% 100	% 100	% 100

المصدر: George Argyrous, Statistics for Social & Health Research, op.cit , p.397.

العينة 1: تصور لنا هذه العينة موقفاً تعكس فيه النسب المشاهدة النسب المتوقعة Expected Percentages، إلى حد بعيد مفترضين أن المتغيرين مستقلان. وفي أحيان أخرى، ربما نجد موقفاً آخر يظهر في العينة الثانية حيث يمكننا ملاحظة وجود تباين كبير بين المجموعات، ولكن هذا التباين ليس بالحد الكبير جداً. في حين تعكس العينة 3: موقفاً متطرفاً نتوقع من خلاله الحصول على حالات من أي طرف من المقياس مسببة التوزيعات النسبية في أول العمودين لتنحرف كثيراً عن تلك التوزيعات النسبية لمجموع العمود. وبالرغم من أن هذه احتمالية عند المعاينة العشوائية من مجتمعات حيث لا توجد علاقة. وهي أيضاً بعيدة الاحتمال.

في الحقيقة أنه باستطاعتنا أن نأخذ عدداً لا متناهياً من العينات العشوائية من مجتمعات يكون فيها المتغيران مستقلين ومشاهدة توزيع النتائج.

وبوضوح فإن معظم النتائج ستكون مشابهة لما لاحظناه في العينتين 1 و 2، وأن قليلاً من النتائج التي تشبه نتائج العينة رقم 3. أن إحصاء مربع كاي يعتبر أداة فعالة، تساعد الباحث في معرفة هذا الفرق بين التكرارات المشاهدة والتكرارات المتوقعة.

ويحسب إحصاء مربع كاي من الفرق بين التكرارات المشاهدة والتكرارات المتوقعة لكل خلية من خلايا الجدول الثنائي Bivariate Table، وأن المعادلة المستخدمة في حساب مربع كاي هي

$$X^2 = \sum \frac{(fo - fe)^2}{fe}$$

حيث إن: F0 تشير للتوزيعات المشاهدة في كل خلية.

Fe: تشير للتوزيعات المتوقعة في كل خلية.

أحياناً يقوم الباحث بسحب عينات تكون "حقيقية" للمجتمع، وبالتالي لا يوجد فرق بين التكرارات الفعلية، والتكرارات المتوقعة. بمعنى آخر، يمكن للباحث أن يتحصل على توزيعات خلية مماثلة لتلك التوزيعات في الجدول رقم 3. في هذه الحالة يتحصل الباحث على قيمة مساوية لصفر لمربع كاي:

$$fo = fe \rightarrow (fo - fe)^2 = 0 \rightarrow X^2 = 0$$

وتجدر الإشارة هنا إلى أنه لن يكون الأمر كذلك في كل عينة. أحياناً سوف يأخذ الباحث عينات من خلال فرصة عشوائية، قد لا تعكس تماماً المجتمعات التي سحبت منها هذه العينات. ومن هنا فإن النتيجة لمربع كاي ستأخذ قيمة موجبة:

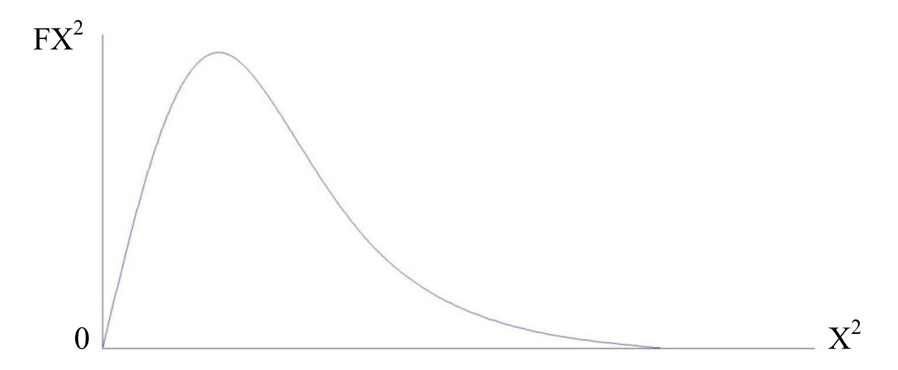
$$fo \neq fe \rightarrow (fo - fe)^2 > 0 \rightarrow X^2 > 0$$

وتجدر الإشارة إلى أنه كلما كان الفرق كبيراً بين التوزيعات التكرارية المشاهدة والتوزيعات التكرارية المتوقعة كانت قيمة مربع كاي كبيرة. لاحظ أنه في معادلة مربع

كاي، أن الفروق بين التوزيعات التكرارية المشاهدة، والتوزيعات التكرارية المتوقعة قد تم تربيعها. ويؤكد هذا التربيع أن المدى لكل القيم المحتملة لمربع كاي يجب أن تبدأ عند صفر وتزداد في الاتجاه الموجب، بغض النظر عما إذا كان التكرار المتوقع أكبر من التكرار المشاهد أو العكس بالعكس. إن تربيع أي فرق سوف يولد رقماً موجباً (بما أنه قد تم حساب مربع كاي على أساس الفرق بين الدرجات المتوقعة والدرجات الفعلية المربعة، وليس على أساس اتجاه الفرق حيث لا جدوى أن نختار بين الاختبار الاستدلالي أحادي الجانب أو ثنائي الجانب. فكل الفروق بين الدرجات المشاهدة والدرجات المتوقعة - بغض النظر عما إذا كانت هذه الفروق قد تكون ناشئة من الدرجات المشاهدة سواء كانت أعلى أو أدنى من الدرجات المتوقعة - سوف تأخذ قيمة موجبة)[2].

إن توزيع مربع كاي يحتوي على ذيل طويل انظر الشكل رقم (1)، الذي يعكس الحقيقة التي مفادها أن هناك احتمالية اختيار عينات عشوائية تعطي قيمة عالية لمربع كاي، بالرغم من أن المتغيرات تكون مستقلة، إلا أن هذا الأمر قد يكون بعيد الاحتمال. إن هذا الأمر سيكون مجرد ضربة حظ كي يختار الباحث عينة من مجموعة واحدة تكون الحالات فيها قد جاءت من نهاية طرف واحد من التوزيع، وإن عينة أخرى من مجموعة أخرى قد جاءت من نهاية الطرف الآخر من التوزيع إذا كان الفرض الصفري للاستقلال صحيحاً. وعليه، فإن المساحة تحت المنحنى للقيم الكبيرة لكاي تربيع تكون صغيرة بحيث تعكس احتمالية منخفضة لما حدث بالصدفة[3].

(لاحظ من خلال توزيع المعاينة لمربع كاي أنه يمكننا تحديد الاحتمالية بأن الفرق بين الدرجات المشاهدة والدرجات المتوقعة تكون ناتجة عن التباين العشوائي عند عملية المعاينة من مجتمعات يكون فيها المتغيران مستقلين). فعلى سبيل المثال، من الممكن أن نجد عينة (3) التي تناولناها سابقاً سوف تسحب مرة واحدة في الألف ($P = 0.001$) إذا كان المتغيران مستقلين عن بعضهما البعض. إن مثل هذا الأمر يمكن التفكير فيه، ولكنه على نحو بعيد الاحتمال كمبرر لنا من أن فرضيتنا حول الاستقلال يجب التخلي عنها - في الواقع ثمة تبعية أو اعتماد dependence بين الدخل ومكان الإقامة.



شكل (18 ـ 1) توزيع X^2

دعنا الآن نتناول المثال المتعلق بالدخل ومكان الإقامة لكي نوضح إجراءات اختبار مربع كاي للاستقلال بشكل عملي. فالمسح النظري الذي تناولناه سابقاً والذي يحتوي على 323 فرداً من المناطق الريفية و 397 فرداً من المناطق الحضرية، مع توزيع الإجابات يوضحه الجدول (18 – 5).

جدول (18 ـ 5) مكان الإقامة حسب الدخل
(التوزيعات المشاهدة)

مكان الإقامة	الدخل		
	منخفض	عال	المجموع
ريفي	205 (%55)	118 (%34)	323 (%45)
حضري	167 (%45)	230 (%66)	397 (%55)
المجموع	372 (%100)	348 (%100)	720 (%100)

نرى من خلاله أن أول رقم في كل خلية يشكل العدد الحقيقي لذوي الدخل المنخفض وذوي الدخل العالي الذين يعيشون في كل منطقة من هاتين المنطقتين، وتمثل 55 % كل الذين سجلوا بأن دخولهم منخفضة، ويعيشون في مناطق ريفية حيث يصل

عددهم إلى 205 فرداً. وعلى الجانب الآخر، فإن نسبة 34 % فقط من ذوي الدخول العالية يعيشون في مناطق ريفية (118 فرداً).

وتجدر الإشارة إلى أنه من خلال هذه البيانات يمكننا أن نرى الفرق بين أولئك الأفراد ذوي الدخول المنخفضة والأفراد ذوي الدخول العالية فيما يتعلق بالمناطق التي يعيشون فيها. إن السؤال الذي يمكن طرحه هنا هو: هل هذا الفرق يكون راجعاً إلى التباين العشوائي؟ للإجابة على هذا السؤال نحتاج، بادئ ذي بدء، إلى حساب التوزيعات المتوقعة: أي أن الأعداد المتوقع الحصول عليها في كل خلية من خلايا الجدول إذا كان المتغيران مستقلين (انظر الجدول 18 – 6).

جدول (18 ـ 6) مكان الإقامة حسب مستوى الدخل
(التوزيعات المتوقعة)

مكان الإقامة	الدخل		
	منخفض	عال	المجموع
ريفي	$\frac{45}{100} \times 372 = 167.4$	$\frac{45}{100} \times 348 = 156.6$	45 %
حضري	$\frac{55}{100} \times 372 = 204.64$	$\frac{55}{100} \times 348 = 191.4$	55 %
المجموع	372	348	100 %

إن عدد المبحوثين المتوقع أن نجدهم في كل خلية إذا كان المتغيران مستقلين، تم حسابهم في كل خلية. فعلى سبيل المثال، إذا كان 45 % لكل المبحوثين يعيشون في مناطق ريفية، فإننا نتوقع أن نجد 45 % من ذوي الدخل المنخفض يعيشون في مناطق ريفية. يوجد لدينا عددٌ إجماليٌّ قدره 372 فرداً من ذوي الدخل المنخفض، 45 % من هؤلاء يعطينا 167 فرداً من ذوي الدخول المنخفضة المتوقع معيشتهم في مناطق ريفية وهكذا في باقي الخلايا (انظر جدول 18 – 7).

جدول (18 ـ 7) مكان الإقامة حسب مستوى الدخل (التوزيعات المشاهدة والتوزيعات المتوقعة)

مكان الإقامة	الدخل		
	منخفض	عال	المجموع
ريفي	205 (167.4)	118 (156.6)	323
حضري	167 (204.6)	230 (191.4)	397
المجموع	372	348	720

يتبين من خلال الجدول أعلاه أن هناك فرقاً بين التوزيعات المشاهدة والتوزيعات التكرارية في كل خلية، وأنه بإمكاننا استخدام معادلة مربع كاي لبيان هذا الفرق في رقم واحد. ويوضح الجدول (18 – 8) إجراء حساب مربع كاي من هذه الفروق.

جدول (18 ـ 8) حساب قيمة مربع كاي (X^2)

مكان الإقامة	الدخل		المجموع
	منخفض	عال	
ريفي	$X^2 = \frac{205 - (167.4)^2}{167.4} = 8.5$	$X^2 = \frac{118 - (156.6)^2}{156.6} = 9.5$	323
حضري	$X^2 = \frac{167 - (204.6)^2}{204.6.4} = 6.9$	$X^2 = \frac{230 - (191.4)^2}{191.4} = 7.4$	397
المجموع	372	348	720

وبعد إجراء عملية حساب قيمة مربع كاي (X^2) لكل خلية من خلايا الجدول، يمكننا بعدئذٍ إضافة هذه القيم بعضها لبعض لنتحصل على القيمة النهائية لمربع كاي وللجدول التقاطع ككل التي تعبر عن الحقيقة بأن العينة الفعلية لا تتطابق تماماً مع الفرض الصفري للاستقلال:

$$x^2 = \sum \frac{(Fo - Fe)^2}{Fe}$$ العينة

$$= \frac{(205-167.4)^2}{167.4} + \frac{(118-156.6)^2}{156.6} + \frac{(167-204.6)^2}{204.6} + \frac{(230-191.4)^2}{191.4}$$
$$= 8.5 + 9.5 + 6.9 + 7.8 = 32.7$$

توزيعات مربع كاي The Distribution of X^2:

بعد أن تحصلنا على قيمة مربع كاي مساوية لـِ 32.7، فالسؤال الذي يمكن طرحه هنا هو ماذا تعني هذه القيمة المحسوبة؟ إن هذه القيمة المحسوبة، في حد ذاتها لا تخبرنا بالشيء الكثير، وبمعزلٍ عن الحقيقة فإن هذه القيمة ليست مساوية لصفر، وعليه فهي تشير إلى أن هناك بعض الاعتماد dependence بين هذين المتغيرين في بيانات هذه العينة. والسؤال المطروح هو ما إذا كانت هذه النتيجة تدعونا إلى رفض الفرض الصفري للاستقلال بالاعتماد على احتمالية القيمة المتحَصَّل عليها لهذه العينة وهي (X^2 = 32.7) من المجتمع حيث إن هذين المتغيرين مستقلان. ولتحديد هذه الاحتمالية يمكننا الإشارة إلى الجدول المتعلق بالقيم الحرجة لتوزيعات مربع كاي (انظر الملحق - جدول رقم 1 / 6). وباستخدام هذا الجدول يمكننا الحصول على القيمة الاحتمالية لمربع كاي آخذين في الاعتبار درجات الحرية. ولحساب درجة الحرية، ينبغي اتباع القاعدة التالية:

$$Df = (r-1)\ (C-1)$$

حيث أن : (r) تشير إلى عدد الصفوف.

: (c) تشير إلى عدد الأعمدة.

وفي جدول 2 × 2 كما هو الحال في المثال الذي بين أيدينا فإن درجة الحرية هي (1). وإن القيمة الحرجة المناظرة هي 3.841، على مستوى الدلالة 0.05:

$$x^2 = 3.841(\alpha = 0.05, df = 3)7.815$$

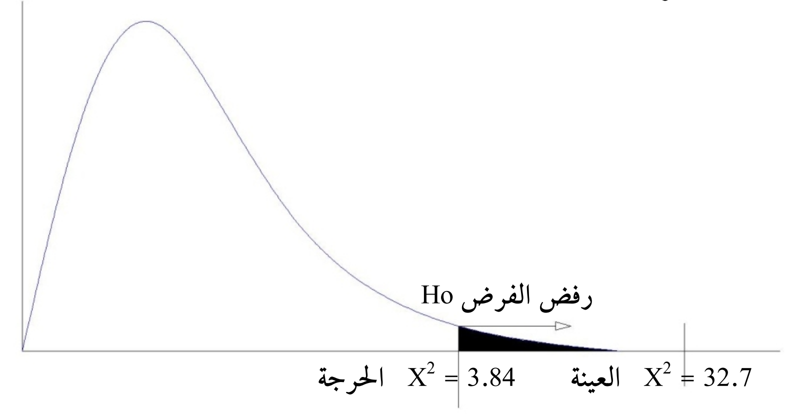
وتجدر الإشارة إلى أنه يمكن رفض الفرض الصفري H0، عندما تكون قيمة X^2 المحسوبة مساوية أو أكبر من قيمة X^2 المتوقعة. في المثال الذي يبين أيدينا، والذي اعتمدنا

قيمة الفا 0.05 بدرجة حرية واحدة، فإن القيمة الاحتمالية لـِ X^2 تساوي 3.841، في حين أن قيمة X^2 المحسوبة للعينة المدروسة هي 32.7. إذاً من خلال هذه المعلومات، يمكننا الوصول إلى نتيجة مفادها: أن القيمة الاحتمالية لمربع كاي تساوي 3.841، وأن القيمة المحسوبة (قيمة العينة) تساوي 32.7، وهي قيمة قد وقعت في المنطقة الحرجة (منطقة الرفض)، وعليه، يمكننا في هذا المثال، رفض الفرض الصفري H0 للاستقلال (مكان الإقامة والدخل ليسا مستقلين).

تجدر الإشارة إلى ملاحظة مهمة، وهي أن اختبار مربع كاي في حد ذاته لا يخبرنا عما تكون عليه طبيعة هذه العلاقة. وأن كل ما يمكن أن نقربه من خلال استخدام هذا الاختبار أن هناك بعضاً من الارتباط بين هذين المتغيرين، وقد اخترنا وصفاً لهذه العلاقة أنها علاقة ذات اتجاه واحد من الدخل إلى مكان الإقامة.

ربما أن شخصاً ما قد يدعي بسهولة أن السببية يمكن أن تتخذ اتجاهاً آخراً، وأن هناك علاقة متبادلةً معتمدةً بين هذين المتغيرين Mutual dependence. إن اختبار مربع كاي لا تقرر لنا مثل هذه القضية، بقدر ما يخبرنا بأن هذين المتغيرين ليسا مستقلين. بمعنى أنهما يرتبطا بعضهما ببعض[4].

في هذا السياق إذاً، كيف لنا أن نختار وصفاً لهذه العلاقة. إن الإجابة على ذلك تعتمد في الأساس على مسألة الجدال النظري الذي يقر بأن التحليل الإحصائي يستطيع أن يخبر ولكنه لا يقرر أبداً[5].



شكل (18 ـ 2) درجات القيم الحرجة والعينة

إجراء اختبار مربع كاي باستخدام برنامج SPSS:

توليد جداول التقاطع Crosstabs لمربع كاي باستخدام Spss:

1- من القائمة الموجودة في الجزء من الشاشة اختر:

Analyze ⟸ Summarize ⟸ Crosstabs

2- انقر على المتغير في القائمة التي تشكل صفوف الجدول في هذه الحالة Place of Residence.

3- انقر على ◄ الذي يشير إلى القائمة المحددة بعنوان Row(S)، تقود هذه العملية إلى لصق (Pastes) Place of Residence في القائمة المحددة لـِ Row(S) .

4- انقر على المتغير في القائمة التي تشكل أعمدة الجدول في هذه الحالة Income Level.

5- انقر على ◄ الذي يشير إلى القائمة المحددة بعنوان Column(s). هذه العملية تقود إلى لصق Income Level في القائمة المحددة لـِ Column(s).

6- انقر على زر Statistics، هذا يقود إلى إعطائنا مربع Crosstabs :Statistics.

7- اختر ch-Square بالنقر على الصندوق القريب منه بوضع علامة T في الصندوق للتأكد بأنه قد تم اختياره.

8- انقر على Continue.

9- انقر على OK.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

Crosstabs

Case Processing Summery

	CASES					
	Valid		Missing		TOTAL	
	N	Percent	N	Percent	N	Percent
Place of Residence Income Level	720	100.0	0	%0	720	%.0100

Place of Residence* Income Level Crosstabulation

	Income Level		TOTAL
	Low	High	
Place of Residence* urban count	167	230	397
within Income Level%	44.9%	66.1%	55.1%
Rural Count	205	118	323
within Income Level%	55.1%	33.9%	44.9
ToTAL Count	372	348	720
within Income Level%	100.0%	100.0%	100.0%

Chi - Square Tests

	Valne	Df	Asymp. sig (2-tailed)	Exact sig (2-tailed)	Exact sig (1-tailed)
Pearson Chi - Square	32.6676[b]	1	.000		
Contiuity Correction[a]	31.816	1	.000		
Likeli hood Ratio	32.965	1	.000		
Fisherś exact test				.000	.000
Linear - by Linear Association	32.622	1	.000		
N. of Valid Cases	720				

.a. Completed Only for a 2X2 table

.b. O Cells (0%) have Expected Count is 156.12

المصدر: .George Argyrous, Statistics for Social & Health Research, op.cit , p.406

شكل (18 ـ 3) مخرجات SPSS لمربع كاي

تفسير مخرجات SPSS لمربع كاي (X^2):

إن الشيء المهم الذي ينبغي على الباحث الاهتمام به من هذه المخرجات هي قيمة Pearson chi-Square الموجودة في المربع الأخير المُعَنْوَن باسم chi-Square tests (اختبار مربع كاي). في الصف الأول المُعَنْوَن بـ Pearson chi-Square تحت قيمة 32.667 التي تشير إلى قيمة X^2 للعينة التي تم حسابها سابقاً يدوياً، بدرجة حرية 1 عمود (3)، ودلالة 000. (عمود 4). وبالرغم من أن درجة الدلالة هذه تعني أقل من ألفا 0.05 في 10.000 للحصول على عينات لهذا التوزيع من المجتمعات التي تكون فيها المتغيرات مستقلة. ومن خلال هذه القيمة الاحتمالية المنخفضة فهي تقودنا إلى رفض الفرض الصفري للاستقلال.

أما إذا كان لديك جدول 2 × 2، وكذلك عندما تكون أي خلية من خلايا التوزيعات النظرية صغيرة جداً، أي أقل من 10 حالات حينئذ على الباحث أن يستخدم القيمة الموجودة في الصف الثاني من مربع chi-Square tests، وهذه القيمة هي قيمة Contiuity Correction التي وضعها yates (تعد هذه القيمة تعويضاً للمبالغة في تقدير اختبار X^2 عند استخدام جدول 2 × 2).

وبالنظر إلى القيمة في عمود (Asymp. Sig. (2.tailed يمكن للباحث أن يتخذ القرار المناسب، فإذا كانت هذه القيمة الموجودة في هذا العمود أكبر من قيمة 0.05، عندئذٍ يمكن للباحث التوصل إلى أن النتيجة ليست دالة والعكس بالعكس.

مثال لزيادة التوضيح:

يمكننا الآن العمل من خلال هذا المثال لزيادة التوضيح مستخدمين الخطوات الخمس المتبعة في إجراء اختبار الفروض. ومن خلال البيانات المبينة في الجدول التالي لعدد 800 طفلاً، فيما يتعلق بعلاقة النوع، وما إذا كان هؤلاء الأطفال يشاهدون الأخبار في الإذاعة المرئية.

جدول (18ـ9) مشاهدة الأطفال للأخبار في الإذاعة المرئية

مشاهدة الأخبار في الإذاعة المرئية	النوع		
	إناث	ذكور	المجموع
نعم	377	363	740 (92 %)
لا	25	35	60 (8 %)
المجموع	402	398	800

المصدر: George Argyrous, Statistics for Social & Health Research, op.cit , p.414.

الخطوة الأولى: صياغة الفرض الصفري والفرض البديل.

H0: النوع ومشاهدة الأخبار المرئية متغيران مستقلان عن بعضهما البعض.

Hi: النوع ومشاهدة الأخبار المرئية متغيران غير مستقلين عن بعضهما البعض.

الخطوة الثانية: اختيار اختبار الدلالة.

سنقوم في هذا المثال بتحليل بيانات عينة تم ترتيبها في جدول ثنائي لمعرفة ما إذا كانت هناك علاقة بين هذين المتغيرين اللذَيْن نجري اختبار X^2 للاستقلال عليهما كاختبار ملائم للاستدلال.

الخطوة الثالثة: حساب درجة العينة:

ولحساب درجة العينة يتعين علينا بادئ ذي بدء حساب التوزيعات المتوقعة استناداً إلى مجموع نسب الأعمدة في الجدول (9).

جدول (18 ـ 10) مشاهدة الأطفال للأخبار المرئية وفقا للنوع

مشاهدة الأخبار في الإذاعة المرئية	النوع		
	إناث	ذكور	المجموع
نعم	371.8	368.2	740
لا	30.2	29.8	60
المجموع	402	398	800

$$x^2 = \sum \frac{(Fo - Fe)^2}{Fe}$$

$$= \frac{(377-371.8)^2}{371.8} + \frac{(363-368.2)^2}{368.2} + \frac{(25-30.2)^2}{30.2} + \frac{(35-29.8)^2}{29.8}$$
$$= 1.9$$

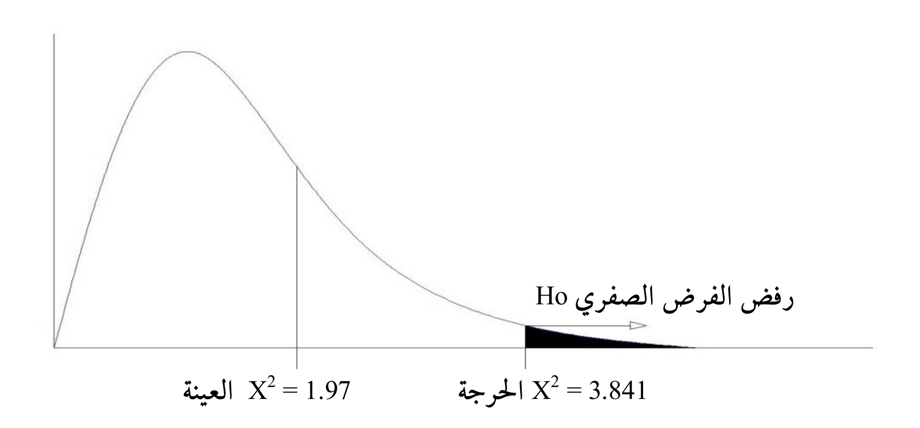
الخطوة الرابعة: إيجاد الدرجة الحرجة والمنطقة الحرجة:

في هذا المثال، نحن نتعامل مع جدول 2 × 2، وعليه، فإننا نتحصل على درجة حرية واحدة على مستوى ألفا 0.05 = α، وبالتالي فإن القيمة الحرجة لِـ X^2 تساوي =

$$x^2 = 3.841(\alpha = 0.05) df = 1$$

الخطوة الخامسة: اتخاذ القرار:

من خلال الشكل (18 – 4) فإن درجة العينة لا تقع في المنطقة الحرجة، وعليه، لا يمكننا رفض الفرض الصفري بأن هذين المتغيرين مستقلان.



شكل (18ـ4) الدرجة الحرجة ودرجة العينة

الافتراضات والقيود في استخدام مربع كاي:

لكي يُسْتَخْدَمَ مربع كاي لجودة التطابق أو اختيار مربع كاي للاستقلال، توجد مجموعة من الشروط ينبغي توفرها. وأن أي اختيار إحصائي يخالف هذه الافتراضات والقيود يبعث بالشك في النتائج التي يتوصل إليها الباحث.

على سبيل المثال، إن الاحتمالية للوقوع في الخطأ من النوع الأول Type I Error يمكن أن يُشَوَّه عندما لا يفي بافتراضات الاختبارات الإحصائية.

هناك بعض الافتراضات والقيود ينبغي مراعاتها عند استخدام اختبارات مربع كاي:

1- يستند مربع كاي على فرضية أن تكون العينات خاضعة للشروط الاحتمالية. ومن هنا يكون من غير المناسب استخدام مربع كاي أو أي مقاييس استدلال لعينات لم يتم حسابها باستخدام العينات الاحتمالية.

2- ينبغي أن تكون المشاهدات مستقلة عن بعضها البعض. في هذه النقطة ينبغي ألا يخلط بينها وبين مفهوم الاستقلالية بين المتغيرات كما رأينا عند الحديث عن اختبار الاستقلال. أن أحد نتائج المشاهدات المستقلة تكون بأن كل تكرار مشاهد يتولد من

خلال موضوع مختلف. بمعنى آخر، أن المشاهدات المستقلة يمكن حسابها من كل فرد أو حالة مرة واحدة فقط، حيث لا يمكن أن يظهر الفرد أو الحالة في أكثر من فئة أو مجموعة. كما أن البيانات المأخوذة من فرد واحد لا يمكن أن تؤثر في المشاهدات المأخوذة من حالة أخرى أو فرد آخر. ومن هنا يكون اختبار مربع كاي اختباراً غير ملائم عندما يكون شخص ما قد أنتج إجابات يمكن تصنيفها في أكثر من فئة أو تسهم في أكثر من حساب توزيع لفئة مفردة بمعنى، أن كل ملاحظة أو قياس لا يكون متأثراً بالملاحظة الأخرى. فالملاحظة أو القياس الذي نتحصل عليه من كل فرد ينبغي إلا يؤثر على البيانات المأخوذة من عينة أو حالة أخرى.

3- حجم التكرارات المتوقعة: تجدر الإشارة إلى أنه ينبغي على الباحث ألا يستخدم اختبار مربع كاي عندما يكون التوزيع التكراري لأي خلية في التوزيع أقل من 5. لأن إحصاء مربع كاي يمكن أن يُشَوَّه عندما يكون التكرار المتوقع صغيراً جداً. انظر على سبيل المثال، لو أنك قمت بحسابات مربع كاي لخلية مفردة. نفرض أن الخلية بها قيمة fe=1 وأن fo=5. فإن حساب هذه الخلية للقيمة الكلية لمربع كاي تكون كالتالي:

$$\text{(الخلية) Cell} = \frac{(f0 - fe)^2}{fe} = \left(\frac{5-1}{1}\right)^2 = \frac{4^2}{1} = 16$$

والآن يمكنك التفكير في مثال آخر تكون فيه Fe= 10 و F0= 14 إن الفرق بين التوزيعات التكرارية المشاهدة، والتوزيعات التكرارية المتوقعة لازال (4)، إلا أن إسهام هذه الخلية للقيمة الكلية لمربع كاي تختلف عن الحالة الأولى:

$$\text{(الخلية) Cell} = \frac{(f0 - fe)^2}{fe} = \frac{(14-10)^2}{10} = \frac{4^2}{10} = 1.6$$

من هنا يتضح جلياً أن قيمة fe صغيرة لديها تأثير كبير على قيمة مربع كاي. وتصبح هذه المشكلة أكثر حدة عندما تكون قيمة fe أقل من 5. وعندما تكون fe صغيرة جداً ستولد قيمة عالية. ومن هنا نجد أن اختبار مربع كاي اختبار حساس عندما تكون قيم fe صغيرة بشكل مفرط. ولتحاشي هذه المواقف، ينبغي على الباحث استخدام عينات كبيرة بدلاً من العينات الصغيرة(6).

قد يصادف الباحث عند تحليل البيانات خلايا مختلفة لديها توزيعات تكرارية أقل من 5. وكنتيجة لذلك، فإن قيمة مربع كاي يمكن أن يكون مغالياً فيها، مما يؤدي إلى رفض الفرض الصفري بالرغم من عدم وجود تطابق بين المتغيرين تحت الدراسة في المجتمع.

إن السؤال المطروح في هذا السياق هو: ماذا ينبغي على الباحث فعله كي يتغلب على هذه المشكلة؟

للتغلب على مثل هذه المشكلة يكون بإمكان الباحث - في بعض الأحيان - اختزال بعض الفئات لواحد من المتغيرين أو كليهما في جدول التوافق لتوليد تكرارات مشاهدة وتكرارات متوقعة كبيرة داخل الخلايا. ومن ناحية ثانية، باستطاعة الباحث أن يجمع فئات المتغيرات، فقط عندما تكون هذه الفئات مرتبطة مع بعضها البعض ارتباطاً منطقياً. وبشكل عام، يمكن للباحث أن يجمع فقط تلك الفئات المتجاورة والمقاسة على متغيرات ترتيبية. فعلى سبيل المثال، يمكن للباحث أن يتعامل مع متغير "زيارة الأقارب"، وذلك بجمع المبحوثين الذين يزورون أقاربهم مرة في العام، مع أولئك الذين أجابوا مرتين أو ثلاث مرات في العام. ولكن ليس من المعقول أن يجمع الباحث إجابات المبحوثين الذين يزورون أقاربهم مرة في الشهر أو مرة في الأسبوع لمجرد أن يحصل الباحث على تكرارات مشاهدة وتكرارات متوقعة كبيرة. كذلك الأمر إذا كان لدى الباحث مجموعات منفصلة مثل: منخفض، منخفض جداً، عال. ففي هذه الحالة يمكن للباحث أن يختزل هذا المقياس إلى فئتين: (منخفض + منخفض جداً) وعال ليصبح المقياس: منخفض وعال.

أما فيما يتعلق بجداول التوافق الصغيرة التي تحتوي على فئتين فقط للمتغير المستقل، والمتغير التابع (جداول 2×2) بدرجة حرية 1 فقط، فإنه ليس بالإمكان استخدام مربع كاي إذا كانت أي من التوزيعات المتوقعة للخلية تحتوي دون 5. أما إذا كانت توزيعات خلية 5 أو أكثر، ولكنها أقل من 10. ففي هذه الحالة يتوجب على الباحث إجراء بعض التعديلات الطفيفة على معادلة مربع كاي، وذلك باستخدام تصحيح ياتس yateś Correction:

$$x^2 = \sum \frac{(|fo - fe| - 0.5)^2}{fe}$$

لاحظ هنا، أننا نطرح 5، من الفرق بين التكرارات المشاهدة والتكرارات المتوقعة لتوزيعات الخلية قبل تربيع النتيجة. ويهدف هذا التعديل إلى التقليل من التأثير على قيمة مربع كاي في الفرق بين التكرارات المشاهدة والتكرارات المتوقعة في كل خلية على حدة.

ثالثاً: من المشكلات الأخرى المرتبطة باستخدام مربع كاي هي تأثر مربع كاي بحجم العينة الكبيرة (مشكلات العينة الكبيرة). فعندما يقوم الباحث بحساب مربع كاي في جدول التوافق مبيناً وجود تطابق ضعيف بين متغيرين في عينة صغيرة، ففي هذه الحالة، فإن قيمة مربع كاي، على الأرجح أنها تشير إلى تطابق يمكن مشاهدته. وأن هذه القيمة المتحصل عليها قد جاءت نتيجة للصدفة. وفي حالة أخرى، عندما يحصل الباحث على تطابق بنفس القوة في عينة كبيرة، فإن قيمة مربع كاي تصبح أكبر. وكنتيجة لذلك، فإن التطابق الذي تم الحصول عليه في العينة الصغيرة، ربما سيؤدي إلى احتمالية قبول الفرض الصفري. أما التطابق الذي تم الحصول عليه في العينة الكبيرة سوف يؤدي إلى احتمالية رفض الفرض الصفري، بالرغم من أن قوة التطابق هي واحدة.

تجدر الإشارة إلى أنه ثمة مَنْ يعتقد من الباحثين أن هذه المشكلة في استخدام مربع كاي تؤدي بهم إلى رفض الفرض الصفري بوجود هذه التطابقات الطفيفة. وكنتيجة لذلك، نلاحظ في كثير من التقارير العلمية أنها تحتوي على تطابق ذي دلالة إحصائية، بالرغم من أن هذا التطابق ضعيف جداً، فإنه من الأرجح أن تكون هذه النتيجة قد جاءت بالصدفة.

في حين يعتقد فريق آخر من الباحثين أن حجم العينة لا تشكل مشكلة في استخدام مربع كاي على الإطلاق. ومن هنا يرى هذا الفريق أنه كلما زاد حجم العينة زادت معه ثقتنا في أن التطابق الذي سنتحصل عليه من هذه العينات سيكون ممثلاً للمجتمع الذي سحبت منه هذه العينات.

إن هؤلاء الباحثين قد ميزوا بين القضايا المتعلقة بالدلالة العملية Practical Significance عن القضايا المتعلقة بالدلالة الإحصائية Statistical Significance. فقد تكون درجة التطابق ذات دلالة إحصائية، لكن ذلك لا يعني البتة أنها ذات دلالة عملية.

ففي العينات الكبيرة حتى درجات التطابق الصغيرة للغاية قد تصل إلى درجة الدلالة الإحصائية. فإذا كانت درجة التطابق قيمتها صغيرة ولها دلالة إحصائية، لكن دلالتها العملية محدودة جداً[7].

اختبار أحادي الجانب واختبار ثنائي الجانب:

إن اختبار أحادي الجانب واختبار ثنائي الجانب الذي تناولناه في متن فصول هذا الكتاب، لا تهمنا كثيراً عند التعامل مع اختبار مربع كاي للاستقلال، وعندما يكون الفرض الذي نود اختباره اتجاهياً، فعلى الباحث استخدام اختبار ثنائي الجانب عند التعامل مع هذا الاختبار.

قياس حجم التأثير لاختبار مربع كاي للاستقلال:

إن اختبار أي فرض مثله مثل اختبار مربع كاي للاستقلال يسعى لتنظيم الدلالة الإحصائية للنتائج التي يتوصل إليها من خلال دراسة ما. وبالتحديد فإن الهدف من وراء أي اختبار هو تحديد ما إذا كانت أنماط أو علاقات المشاهدة في بيانات العينة يرجح أن تكون قد حدثت بالصدفة؛ وهذا يعني بدون أي أنماط متشابهة أو علاقات في المجتمع. أن اختبارات الدلالة عادة لا تتأثر فقط بالحجم، أو قوة تأثيرات المعالجة، ولكنها تتأثر أيضاً بحجم العينات. وكنتيجة لذلك، فإن وجود أي تأثير بسيط يمكن أن يكون ذا دلالة إحصائية إذا ما تَمَّتْ ملاحظتُهُ من خلال عينة كبيرة جداً. وحيث إن تأثير الدلالة ليس بالضرورة أن يعني تأثيراً كبيراً، فإنه بشكل عام. يُنْصَحُ بأَنَّ نتائج اختبار فرض ما يجب أن يصاحبه قياس حجم التأثير. إن هذه النصيحة العامة، يجب أيضاً أن تطبق على اختبار مربع كاي للاستقلال.

معامل فاي ϕ و Cramer's V:

يعتبر معامل فاي ϕ مقياساً للارتباط للجدول الثنائي للبيانات التي تحتوي على متغيرين ثنائيين (كلا المتغيرين يأخذان قيمتين فقط مثل: النوع، ذكر / أنثى... الخ)،

وتنطبق هذه الحالة أيضاً عند إجراء اختبار مربع كاي لمصفوفة تحتوي على 2×2 (مرة أخرى كلا المتغيرين يأخذان قيمتين فقط). في مثل هذه الأحوال يمكننا حساب معامل ϕ، بالإضافة إلى حساب اختبار مربع كاي للاستقلال لنفس البيانات. ولما كانت معامل فاي ϕ مقياساً لقوة العلاقة بدلاً من مقياس الدلالة (تستخدم المعادلة التالية لاختبار دلالة معامل فاي ϕ : $(X^2 = N(\phi^2))$، وبالتالي فهي تقدم لنا مقياس حجم التأثير.

وتجدر الإشارة إلى أن قيمة معامل فاي يمكن حسابها مباشرة من مربع كاي وفقاً للمعادلة التالية:

$$\phi = \sqrt{\frac{x^2}{N}}$$

وتُحَدَّد قيمة معامل فاي بشكل كلي من خلال النسب في مصفوفة البيانات 2×2؛ وهي بالتالي مستقلة تماماً عن الحجم المطلق للتكرارات.

إن قيمة مربع كاي، على أي حال، تتأثر بالنسب وكذلك حجم التكرارات.. إن هذا الفرق يمكن البرهنة عليه من خلال المثال التالي:

مثال:

البيانات التالية تبين توزيع تكراري يقيم العلاقة بين النوع والاتجاه نحو تنظيم الأسرة:

الاتجاه	النوع	
	ذكر	أنثى
أوافق	5	10
لا أوافق	10	5

لاحظ أن هذه البيانات تظهر أن الإناث تفضل تنظيم الأسرة أكثر منه لدى الذكور. ومن خلال هذه البيانات يمكننا الحصول على قيمة مربع كاي مساوية لـِ 3.33 (وهي قيمة غير دالة)، وكذلك قيمة معامل فاي $\phi = 0.3332$. وفي الخطوة التالية نحتفظ بنفس

النسبة تماماً لهذه البيانات ولكننا سنضاعف كل التكرارات. وتظهر هذه العملية في المصفوفة التالية:

الاتجاه	النوع	
	ذكر	أنثى
أوافق	10	20
لا أوافق	20	10

مرة أخرى، نجد أن الإناث يفضلن تنظيم الأسرة، إذا ما قورن ذلك بالذكور. الآن، يمكننا ملاحظة أن العينة تحتوي على 30 من الإناث و30 من الذكور. ويمكننا الحصول على قيمة مربع كاي وفاي من هذه البيانات الجديدة: قيمة مربع كاي تساوي 6.66 وهي قيمة ضعف قيمة مربع كاي في المرة الأولى (وهي الآن قيمة ذات دلالة على مستوى ألفا $0.05 = \alpha$). ولكن قيمة فاي بقيت كما هي دون تغير 0.3332.

ولما كانت النسبة متساوية في كلتا العينتين، فإن قيمة معامل فاي لم تتغير. وعليه يمكننا القول بأنه كلما كانت العينة كبيرة، تقدم أكثر حدث مقنع إذا ما قورنت بالعينة الصغيرة، وأكثر ترجيحاً لتمدنا بنتيجة دالة.

ولتفسير قيمة فاي يمكن للباحث أن يتبع نفس المعيار المستخدم في تقييم الارتباط. وتعتبر قيمة $0.10 = \phi$ (تأثير متوسط) و 0.30 تأثير متوسط و 0.50 تأثير كبير. في بعض الأحيان تربع قيمة ϕ (ϕ^2) لتبين لنا نسبة التباين الذي يمكن وضعه في الاعتبار، كما هو الحال عند التعامل مع قيمة r^2. أما إذا كان اختبار مربع كاي يتعلق بمصفوفة تحتوي على أكثر من 2×2، فإن تعديلاً لمعامل فاي ينبغي إجراؤه. ويُعرف هذا التعديل بـِ: Cramerś V لقياس حجم التأثير:

$$V = \sqrt{\frac{X^2}{n(df_*)}}$$

لاحظ معادلة Cramerś V مطابقة لمعادلة معامل فاي، باستثناء الإضافة المتعلقة بـِ (df*) في مقام المعادلة.

إن قيمة df* ليست مساوية لدرجة الحرية df لاختبار مربع كاي للاستقلال حيث إن df = (C-1) (R-1). وتشير R إلى عدد الصفوف، و C إلى عدد الأعمدة. أما فيما يتعلق بـِ: Cramerś V، فإن قيمة df* تشير إلى أيهما أصغر (R-1) أو (C-1). العَلاَّمة كوهين Cohen في عام 1988 قد اقترح معايير لتفسير معامل Cramerś V كما يوضحه الجدول التالي:

جدول (18 ـ 11): معايير تفسير معامل Cramerś V

حجم التأثير			
درجة الحرية	تأثير بسيط	تأثير متوسط	تأثير كبير
df*=1	0.10	0.30	0.50
df*=2	0.07	0.21	0.35
df*=3	0.06	0.17	0.29

المصدر: J. Cohen , Statistical Power analysis for behavioral Sciences, SAN Diego, CA: Academic Press , 1988.

لاحظ عندما تكون df*=1 كما هو الحال في مصفوفة 2×2، فإن معيار تفسير V بالضبط، هو مساوٍ لمعيار تفسير الارتباط المألوف أو معامل فاي. وللتدليل على ذلك، يمكننا الاستعانة بالبيانات التالية التي تقيم العلاقة بين النوع والرغبة في الاستفادة من خدمات الصحة النفسية.

جدول (18 ـ 12): الرغبة في الاستفادة من خدمات الصحة النفسية

النوع	الاحتمالية نعم	ممكن	الاحتمالية لا	مج
ذكور	11	32	17	60
إناث	34	43	13	90
مج	45	75	30	N=150
$x^2 = 8.23\ df = 2\ (5.9\square)\ \alpha = .05$				

المصدر: Frederick J. Gravetter and Larry B. Wallnau , Statistics for the behavioral Sciences , . op..cit , P. 623

لدينا مصفوفة 2×3 بعدد إجمالي 150 = N مشارك. وقد أنتجت هذه البيانات مربع كاي يساوي 238. = x^2، وباستخدامنا هذه القيمة نتحصل على:

$$V = \sqrt{\frac{x^2}{ndf_*}} = \sqrt{\frac{8.23}{150(1)}} = \sqrt{0.055}$$

$$= 0.23$$

وطبقاً لمعايير كوهين، فإن هذه القيمة المتحصل عليها تشير إلى علاقة صغيرة[8].

القوة Power:

يمكن تحديد القوة من الجدول التالي مستخدمين معلومات حجم التأثير المتمثلة في الحجم الكلي لدرجة الحرية TOTAL df وحجم العينة (N) ('TOTAL 'N)، إن النتيجة التي تحصلنا عليها في المثال السابق أشارت إلى تأثير متوسط الحجم بدرجة حرية مساوية (1). (df = 1). وبالرغم من أن الجدول يحتوي على 200 = N فقط. وحتى هذه القيمة لِـ (N)، فإن القوة تصل إلى 97. بحجم كلي لدرجة الحرية يساوي (2) (للتعامل مع جدول القوة المتوقعة يستخدم الحجم الإجمالي لدرجة الحرية).

من الواضح أن القوة في هذا المثال حقاً عالية جداً، وبالطبع فهي قوة مقبولة. وتعني القوة في هذا المثال، أننا مفترضون أن الفرض الصفري فرضُ صحيحُ. إن فرصة 3% تقع

إذا ما أعيد بنفس الحجم (N). ولما كانت حسابات حجم التأثير والقوة حسابات علمية دقيقة فإن الباحث يمكنه الرجوع إلى جداول كوهين التفصيلية[9].

جدول (18-13): قوة اختبار مربع كاي X^2 على مستوى دلالة α= 0.05 و 0.01 α بدرجة حرية 1، 2، 3، 4، 5

درجة الحرية df	الحجم الكلي للعينة (N)	حجم التأثير α = 0.05			حجم التأثير α = 0.01		
		كبير	متوسط	صغير	كبير	متوسط	صغير
1	25	.70	.32	.08	.47	.14	.02
	50	.94	.56	.11	.83	.32	.03
	100	❋	.85	.17	.99	.66	.06
	200	❋	.99	.29	❋	.95	.12
2	25	.60	.25	.07	.36	.10	.02
	50	.90	.46	.09	.74	.24	.02
	100	❋	.77	.13	.98	.55	.04
	200	❋	.97	.23	❋	.91	.08
3	25	.54	.21	.07	.30	.08	.01
	50	.86	.40	.08	.68	.19	.02
	100	.99	.71	.12	.97	.48	.03
	200	❋	.96	.19	❋	.87	.07
4	25	.50	.19	.06	.26	.07	.01
	50	.82	.36	.08	.62	.16	.02
	100	.99	.66	.11	.96	.43	.03
	200	❋	.94	.17	❋	.84	.06
5	25	.45	.17	.06	.23	.06	.01
	50	.79	.33	.07	.58	.14	.02
	100	.98	.62	.10	.94	.38	.03
	200	❋	.93	.16	❋	.80	.05

❋ الدرجة قريبة جداً من 1 صحيح.

المصدر: J. Cohen , Statistical Power analysis for the behavioral Sciences ,SAN Diego , CA: Academic Press , 1988 , PP. 228 - 230 and PP. 235 - 237.

تطبيقات خاصة لاختبارات مربع كاي: [10]

في بداية هذين الفصلين قد قدمنا اختبارات مربع كاي كأمثلة للاختبارات غير البارامترية. وبالرغم من أن الاختبارات غير البارامترية اختبارات تؤدي وظيفة فريدة خاصة بهذه الاختبارات، إلا أنه يمكن النظر إلى هذه الاختبارات كبديل لتقنيات الاختبارات البارامترية (المعلمية) الشائعة التي تم مناقشتها في متن هذا الكتاب.

وبشكل عام، تستخدم الاختبارات غير البارامترية كبديل للتقنيات المعلمية في مواقف قد يحدث في واحد من هذين الموقفين:

1- إن البيانات المجمعة قد لا تلبي الافتراضات المطلوبة لخصائص الاختبار المعلمي.

2- قد تحتوي البيانات على مقياس اسمي أو مقياس ترتيبي قد يصعب معه حساب معلمات المجتمع: كالمتوسط الحسابي، والانحراف المعياري.

في هذا الجزء، سنتناول بعض العلاقات بين اختبارات مربع كاي، والإجراءات البارامترية التي يمكن أن يُسْتَخْدَمَ فيها اختبار مربع كاي كبديل لهذه الإجراءات.

أ- مربع كاي وارتباط بيرسون (r).

ب- مربع كاي والمقاييس المستقلة لِـ وأنوفا ANOVA.

ج- معامل فاي ϕ.

د- اختبار الوسيط للعينات المستقلة.

أـ مربع كاي وارتباط بيرسون (r):

إن اختبار مربع كاي للاستقلال ومعامل ارتباط بيرسون (r) كلاهما تقنيات إحصائية تهدفان إلى تقييم العلاقة بين متغيرين. ويحدد نوع البيانات المتحصل عليها في دراسة معينة أي من الإجراءات الإحصائية لهذين النمطين يمكن أن يكونا ملائماً.

نفترض على سبيل المثال، أن باحثاً يرغب في استقصاء العلاقة بين تقدير الذات Self-Esteem والتحصيل الأكاديمي Academic Performance لعدد من الأطفال يبلغون من العمر عشر سنوات. فإذا تحصل الباحث على درجات عددية Numerical Scores لكلا المتغيرين، فإن نتائج البيانات ستكون مشابهة للقيم التالية:

المشاركون في الدراسة	تقدير الذات (X)	الأداء الأكاديمي
1	13	73
2	19	88
3	10	71
4	22	96
5	20	90
6	15	82
:	:	:
:	:	:

المصدر: Frederick J. Gravetter and Larry B. Wallnau , op. Cit , P. 630

وأن الباحث يمكنه أن يستخدم معامل ارتباط بيرسون (r) من أجل تقييم هذه العلاقة. وعلى الجانب الآخر، يمكن للباحث أن يختار، ببساطة، أن يصنف الأفراد إلى فئات لكلا المتغيرين. فعلى سبيل المثال، يمكن له أن يصنف كل طالب، إما إلى تحصيل عال أو تحصيل منخفض. وتصنيف متغير تقدير الذات إلى: تقدير عال، متوسط، ومنخفض.

إن نتيجة البيانات ستولد توزيعاً تكرارياً يمكن ملاحظته في مصفوفة مثل تلك المصفوفة الواردة في الجدول التالي:

جدول (18 ـ 14): توزيع تكراري لمستوى تقدير الذات وفقا للأداء الأكاديمي لعينة من 150 طفلاً يبلغون من العمر عشر سنوات

	مستوى تقدير الذات			
الأداء الأكاديمي	عال	متوسط	منخفض	مج (N)
عال	17	32	11	60
منخفض	13	43	34	90
مج (N)	30	75	45	N=150

المصدر: Ibid , P. 631.

لاحظ أن هذه البيانات لا تتضمن أي درجات عددية. ولكنها تحتوي مجموعة تكرارات تتلاءم واختبار مربع كاي.

بـ مربع كاي والمقاييس المستقلة t وأنوفا:

مرة ثانية، يمكننا النظر إلى باحث يود استقصاء العلاقة بين تقدير الذات، والأداء الأكاديمي لمجموعتين من الأطفال يبلغون من العمر عشر سنوات. نفترض في هذه المرة، أن الباحث قام بقياس الأداء الأكاديمي بتصنيف الأفراد إلى فئتين (عال ومنخفض)، وبعد ذلك تحصل عل درجة عددية لتقدير الذات لكل طفل. وأن نتيجة البيانات ستكون مشابهة للدرجات الواردة في جدول (أ). وأن اختبار t للعينات المستقلة سيستخدم لتقييم فرق المتوسط بين المجموعتين من الدرجات. وكخيار بديل، يمكن للباحث أن يقيس تقدير الذات بتصنيف الأفراد إلى ثلاث فئات: عال، متوسط، منخفض. وإذا تحصل على درجة عددية للأداء الأكاديمي لكل فرد، فإن نتيجة البيانات تظهر في شكل الدرجات في الجدول (ب)، وأن اختبار أنوفا يمكن إجراؤه لتقييم الفروق في المتوسطات بين المجموعات الثلاثة.

جدول (18 ـ 15)

البيانات المناسبة للمقاييس المستقلة لاختبار t أو لاختبار أنوفا الجزء (أ) درجات تقدير الذات تم الحصول عليها من مجموعتين من الطلاب مختلفين في مستوى الأداء الأكاديمي. والجزء (ب) درجات الأداء الأكاديمي التي تم الحصول عليها لعدد ثلاث مجموعات مختلفين في مستوى تقدير الذات

(أ) :

درجات تقدير الذات لمجموعتين من الطلاب الأداء الأكاديمي	
عال	منخفض
17	13
21	15
16	14
24	20
18	17
15	14
19	12
20	19
18	16

(ب):

درجات الأداء الأكاديمي لثلاث مجموعات تقدير الذات		
عال	متوسط	منخفض
94	83	80
90	76	72
85	70	81
84	81	71
89	78	77
96	88	70
91	83	78
85	80	72
88	82	75

المصدر: Ibid , P. 631.

إن الهدف من وراء بيان هذه الأمثلة، هو أن اختبار مربع كاي للاستقلال، ومعامل ارتباط بيرسون (r)، واختبارات فروق المتوسط جميعها تقنيات إحصائية يمكن توظيفها لتقييم العلاقة بين متغيرين. إن التمييز الأساسي بين مختلف هذه الإجراءات الإحصائية يكمن في شكل البيانات. ومع ذلك، هناك تميزُ آخر هو الهدف الأساسي لهذه الإحصاءات المختلفة.

إن اختبار مربع كاي، واختبارات فروق المتوسط (t، وأنوفا) تستخدمان لتقييم دلالة

العلاقة، بمعنى أن هذه التقنيات تقرر ما إذا كانت العلاقة المشاهدة في العينة تمدنا بدليل كافي لنقرر أن هناك علاقة تشابه في المجتمع. وكذلك يمكن للباحث أن يقيم دلالة ارتباط بيرسون (r)، ومع ذلك، فإن الهدف الأساسي للارتباط هو قياس قوة Strength العلاقة. بالخصوص، إن تربيع قيمة r^2، تمدنا بقياس حجم التأثير Effect size وذلك بوصف نسبة التباين Proportion of Variance في أحد المتغيرين وما يحدثه في علاقته مع المتغير الآخر.

ج- معامل فاي ϕ :

لقد بينا في الفصول السابقة أن العلاقة بين متغيرين ثنائيين يمكن تقييمها من خلال، إما معامل فاي ϕ أو اختبار مربع كاي للاستقلال لمصفوفة 2×2. ففي المثال الذي تناولناه في الصفحات السابقة، واختبرنا فيه العلاقة بين النوع (ذكور/ إناث) وتفضيلها فريق أ أو ب. نلاحظ أن معامل فاي ϕ أنتجت ارتباطاً من شأنه أن يقيس قوة العلاقة. في حين أن اختبار مربع كاي يقيم دلالة هذه العلاقة.

د- اختبار الوسيط للعينات المستقلة:

يمدنا اختبار الوسيط بالإحصاءات غير البارامترية بديلة للمقاييس المستقلة لـِ t أو أنوفا ANOVA لتحديد ما إذا كان هناك فروق دالة بين اثنين أو أكثر من العينات المستقلة. ويصاغ الفرض الصفري لاختبار الوسيط على أن عينات عدة جاءت من مجتمعات تتقاسم وسيطاً مشتركاً (لا توجد فروق). في حين يصاغ الفرض البديل على أن العينات جاءت من مجتمعات عدة لا تتقاسم وسيطاً مشتركاً.

إن المنطق وراء اختبار الوسيط هو أنه متى يتم اختيار عدة عينات مختلفة من نفس توزيع المجتمع، فإن نصف الدرجات تقريباً في كل عينة ينبغي أن تكون هذه الدرجات فوق وسيط المجتمع، والنصف الآخر من الدرجات ينبغي أن يكون تحت الوسيط. بمعنى أن كل العينات المنفصلة ينبغي أن تكون موزعة حول نفس الوسيط. وفي الجانب الآخر، إذا كانت العينات قد جاءت من مجتمعات لديها وسيطات مختلفة، حينئذٍ فإن الدرجات في بعض العينات سوف تكون وبشكل متساوق أعلى من الدرجات في عينات أخرى، وبالتالي ستكون بشكل مُتَّسِقٍ أقل.

إن أولى الخطوات لإجراء اختبار الوسيط هو دمج كل الدرجات من العينات المنفصلة، وبعد ذلك، إيجاد الوسيط لهذه المجموعات المندمجة. والخطوة الأخرى، هي بناء مصفوفة Matrix تحتوي على عمود لكل عينة من العينات، وصفين، وتحسب درجة للأفراد الذين هم فوق الوسيط، ودرجة للأفراد الذين هم تحت الوسيط. وأخيراً، يتم عد كم عدد الأفراد الذين سجلوا درجات فوق الوسيطات المشتركة Combined Median، وكم عدد الأفراد الذين سجلوا درجات تحت الوسيط. وبعد ذلك تنظم هذه التكرارات المشاهدة في مصفوفة. ويمكن تقييم هذه التكرارات المشاهدة من خلال استخدام مربع كاي للاستقلال.

إن التوزيعات التكرارية وقيمة مربع كاي يمكن حسابهما بالطريقة نفسها التي تم توضيحها في هذا الجزء من هذا الكتاب. إن القيمة الدالة لاختبار مربع كاي تشير إلى أن التعارض بين توزيع أفراد العينة الأولى يكون أكبر مما هو متوقع عن طريق الصدفة. ولتوضيح اختبار الوسيط للعينات المستقلة نورد المثال التالي:

مثال: البيانات التالية تبين درجات تقدير الذات تم الحصول عليها من عينة تتألف من أربعين طفلاً (N=40). وبعد ذلك تم تقسيم هؤلاء الأطفال إلى ثلاث مجموعات منفصلة طبقاً لمستوى الأداء الأكاديمي (عال، متوسط، منخفض). وباستخدام الوسيط يمكن للباحث أن يقيم ما إذا كانت العلاقة بين هذين المتغيرين علاقة دالة.

جدول (18 ـ 10)

درجات تقدير الذات للأطفال وفقاً لثلاثة مستويات من الأداء الأكاديمي [11]

منخفض			متوسط				عال			
	19	11	20	24	13	22		14	22	
- 7	15	13	16	10	22	18	+ 9	18	19	+ 8
+ 3	16	20	19	14	15	19	– 11	21	12	– 2
= 10	18	10	10	11	18	11	= 20	20	18	= 10
	11	15	12	15	19	12		23	20	

المصدر: Ibid , P. 632.

ولأجل اختبار هذه الفرضية ينبغي على الباحث أن يقوم بالخطوات التالية:

- دمج المجموعات الثلاثة لتصبح كما لو كانت هذه المجموعات مجموعة واحدة.

$$N = 40 = 10 + 20 + 10 = N_3 + N_2 + N_1$$

- يتم ترتيب درجات العينة الجديدة المدمجة تصاعدياً أي من أصغر درجة وهي 10 وحتى أكبر درجة وهي 24. ثم يستخرج الوسيط لهذه المجموعة من الدرجات بنفس الإجراءات التي بيناها في أحد مواضع هذا الكتاب.

إن الوسيط لهذه المجموعة المدمجة N = 40 هو M=16+18=34/2=17. أي 20 درجة تماماً فوق هذه القيمة الوسيطية، وعشرون درجة تحت الوسيط. ونجد أن 8 درجات من أصل 10 درجات فوق الوسيط عندما يتعلق الأمر بالأداء الأكاديمي العالي. في حين نجد 9 درجات من أصل 20 درجة فوق الوسيط فيما يتعلق بالأداء المتوسط. بينما نجد أن ثلاث درجات فقط من أصل عشر درجات فوق المتوسط عند الحديث عن الأداء الأكاديمي المنخفض. إن هذه التوزيعات المشاهدة يمكن تنظيمها في المصفوفة التالية:

الأداء الأكاديمي			
	عال	متوسط	منخفض
فوق الوسيط	8	9	3
تحت الوسيط	2	11	7

والجدول التالي يوضح التوزيعات التكرارية لهذا الاختبار:

الأداء الأكاديمي			
	عال	متوسط	منخفض
فوق الوسيط	5	10	5
تحت الوسيط	5	10	5

من خلال هذا الجدول يمكننا استخراج قيمة كاي المحسوبة:

$$x^2 = \frac{(8-5)^2}{5} + \frac{(2-5)^2}{5} + \frac{(9-10)^2}{10} + \frac{(11-10)^2}{10} + \frac{(3-5)^2}{5} + \frac{(7-5)^2}{5}$$

$$x^2 = 5.40$$

بعد إيجاد قيمة مربع كاي المحسوبة، نقوم باستخراج قيمة كاي الجَدْوِلِيَّة ومقدارها 5.99 بدرجة حرية (2) وعند مستوى دلالة 0.05 = α. ولما كانت الدرجة المحسوبة لمربع كاي تساوي 5.40، فإنها لا تقع في المنطقة الحرجة، وعليه لا يمكننا رفض الفرض الصفري (Ho) باعتبار أن هذه البيانات لا تقدم لنا وقائع كافية لنصل إلى نتيجة مفادها أنه لا توجد فروق دالة بين توزيعات تقدير الذات لهذه المجموعات الثلاث من التلاميذ.

وأخيراً نود أن نضمن بعض الملاحظات المتعلقة بتفسير اختيار الوسيط، يمكن إجمالها في الآتي:

أولاً: إن اختبار الوسيط ليس اختباراً للفروق بين المتوسطات، باعتبار أن المتوسط الحسابي لأي توزيع يمكن أن يتأثر بشكل قوي بالدرجات المتطرفة. وعليه، فإن المتوسط والوسيط لأي توزيع ليس بالضرورة أن يكونا متساويين، وأنه ليس من الممكن أن يكونا مرتبطين، إن النتيجة المتحصل عليها من خلال اختبار الوسيط لا نستطيع تفسيرها كما يشار إلى أنه يوجد أو لا يوجد فرق بين المتوسطات.

ثانياً: إن اختبار الوسيط، كما أشرنا، لا يقارن مباشرة الوسيط من عينة واحدة بالوسيط للعينة الأخرى. وبالتالي، فإن اختبار الوسيط ليس اختباراً لدلالة الفروق بين الوسيطات. ولكنه بدلاً من ذلك فهو اختبار يقارن توزيع الدرجات لعينة واحدة في مقابل توزيع الدرجات في عينة أخرى. فإذا كانت العينات قد وزعت بشكل متساوٍ حول نقطة مشتركة (وسيط المجموعة المدمجة)، فإن الاختبار سوف يتضمن أنه لا يوجد فرق ذو دلالة. وعلى الجانب الآخر، فإن إيجاد فرق دال، يشير ببساطة إلى أن العينات ليست موزعة بشكل متساوٍ حول الوسيط المشترك، وبالتالي، فإن أفضل تفسير للنتيجة الدالة هو أنه يوجد فرق في توزيعات العينات[12].

أسئلة للمراجعة :

1- ما هي درجة الحرية للجداول ذات الأبعاد التالية:

– 2×2   – 6×4   4×2   – 3×3

2- ما هي القيم الحرجة لـِ X^2 بدلالة 0.05 و 0.10 = ∝ للجداول ذات الأبعاد التالية:

– 2×4   – 6×4   4×2   – 3×3

3- إذا كان X^2 لعينة يساوي 500 أنتجت قيمة لمربع كاي تساوي 24، هل X^2 بنفس التوزيع ذي الصلة ولكن مع:

N=50 , N=100

4- من الجدول التالي – احسب التوزيعات المتوقعة لكل خلية مبيناً تلك الحالات التي تنتهك القواعد المستخدمة في X^2:

	أ	ب	ج	د	المجموع
أ	1	0	6	48	55
ب	2	0	7	40	49
المجموع	3	0	13	88	104

5- في الفصول السابقة قارنا عينات افتراضية للأطفال من استراليا، كندا، سنغافورة، وبريطانيا، فيما يتعلق بكمية مشاهدة الإذاعة المرئية (التلفزيون). افترض أن هذا المتغير لم يتم قياسه على مستوى المقياس الترتيبي والنسبي، ولكن عوضاً عن ذلك تم قياسه على مستوى المقياس الترتيبي. وقد جاءت نتائج هذا المسح كالتالي:

كمية مشاهدة الإذاعة المرئية	المجتمعات				
	كندا	أستراليا	بريطانيا	سنغافورة	المجموع
منخفضة	23	25	28	28	104
متوسطة	32	34	39	33	138
عالية	28	30	40	35	133
المجموع	83	89	107	96	375

من خلال هذه البيانات، هل يمكنك القول بأن كمية مشاهدة الأطفال للإذاعة المرئية مستقلة عن مكان الإقامة؟

6- عينة مكونة من 162 رجلاً تتراوح أعمارهم بين 40 و 65 سنة، تم سحبها لمعرفة أوضاع هؤلاء الرجال الصحية. تم طرح سؤال على كل رجل لمعرفة فيما إذا كانت عادة التدخين على أساس منظم. وقد تم الحصول على النتائج التالية كما تظهر في جدول التقاطع:

الوضع الصحي	عادة التدخين		
	لا يدخن	يدخن	المجموع
سيء	13	34	47
متوسط	22	19	41
جيد	35	09	44
جيد جداً	27	03	30
المجموع	97	65	162

المطلوب:

بالنظر إلى نسب العمود، هل تعتقد بأن الفروق في مستوى الوضع الصحي لعينة المدخنين وغير المدخنين يمكن إرجاعها إلى تباين المعاينة بدلاً من الفرق في المجتمعات.

- أَجْرِ اختبار مربع كاي للاستقلال لهذه البيانات؟ هل النتيجة التي توصلت إليها تؤكد ما توصلت إليه في الإجابة السابقة؟

7- من البيانات التالية:

	الاتجاه		
	أوافق	لا أوافق	
ذكور	35	15	50
إناث	55	45	100
	90	60	

أ- بين ما إذا كان هناك فرق ذو دلالة في توزيع الاتجاهات للذكور مقارنة بالإناث.

اختبر هذا الفرض على مستوى دلالة 05.

ب- إن العلاقة بين النوع والاتجاه يمكن الحصول عليها باستخدام معامل ϕ. ما هي قيمة ϕ إذا تم حسابها من هذه البيانات.

8- البيانات السابقة بينت لنا أنه لا يوجد فرق ذو دلالة بين توزيعات الذكور والإناث. ولبناء جدول جديد يمكنك مضاعفة حجم العينة بـِ 15؛ وبالتالي ستكون التوزيعات لكل خلية ضعفين (لاحظ أن تناسب العينة لم يتغير).

	الاتجاه		
	أوافق	لا أوافق	
ذكور	70	30	100
إناث	110	90	200
	180	120	300

المطلوب:

أ- اختبار فرق الدلالة بين توزيعات الذكور وتوزيعات الإناث مستخدماً مستوى الدلالة 05.، كيف تقارن بين النتائج التي توصلت إليها في التمرين السابق، بما توصلت إليه في التمرين الحالي سوف نجد أنه كلما كبر حجم العينة زاد من احتمالية نتيجة الدلالة.

ب- احسب قيمة ϕ لهذه البيانات، وقارن بين نتيجة ϕ في التمرين السابق، وقيمة ϕ في

التمرين الحالي (نعتقد أنك ستجد أن حجم العينة ليس له أي تأثير على قوة العلاقة).

9- بيانات من قسم المرور بمدينة بنغازي (بيانات افتراضية وليست واقعية) تشير إلى أن 80 % من أولئك الذين يحملون رخص قيادة هم أكبر من 25 سنة. وفي عينة مؤلفة من 50 شخصاً تحصلوا حديثاً على مخالفات مرورية 32 كانوا أكبر من 25 سنة و18 كانوا أعمارهم 25 سنة أو أقل.

السؤال المطروح هو: هل توزيع العمر لهذه العينة يختلف بشكل دال عن التوزيع المتعلق بأولئك الأفراد الذين يحملون رخصاً للقيادة؟ يرجى استخدام $\alpha = .05$.

10- دراسة اجتماعية تهدف إلى تقييم ثلاثة أنواع من أجهزة الهواتف الجوالة وأن الباحث الذي يقوم بهذه الدراسة كانت لديه شكوك أن طلاب الجامعة قد تكون لديهم معايير تختلف عن تلك المعايير لدى الجيل الكبير. ولكي يختبر هذه الفرضية. قد أعد الباحث هذه الدراسة مستخدماً عدد 60 فرداً من الجيل الكبير، إضافة إلى عينة مكونة من 60 طالباً والبيانات التالية تبين نتائج هذه الدراسة.

	تصميم (1)	تصميم (2)	تصميم (3)
الطلاب	27	20	13
الجيل الكبير	21	34	5
مج	48	54	18

السؤال: هل هذه البيانات تشير إلى أن أفضلية التوزيع للجيل الكبير تختلف بشكل دال عن التوزيع لدى طلاب الجامعة.. يرجى اختبار هذه الفرضية بمستوى دلالة $.05\ \alpha$.

11- إذا كانت لدينا قيمة فاي تساوي 0.229 + لجدول 2×2 جاءت من عينة قوامها 200. اختبر ما إذا كانت هذه القيمة دالة؟ (اختبر هذه القيمة على مستوى دلالة 0.05 و 0.01).

الهوامش والمصادر:

أولا: الهوامش:

1- J. Richard Kendrick , Social Statistics: An Introduction using SPSS , Second edition , USA , 2005 , P. 353.

2- George Argyrous , Statistics for Social and Health Research , Sage Publications ,London , 2001 , P. 398.

3- Ibid , P. 399.

4- Ibid , P. 403.

5- Ibid , P. 403.

6- Frederick J. Gravetter and Larry B. Wallnau , Statistics for The Behavioral Sciencies , 8th ed , 2010 , PP. 628 - 629.

7- جولي بالانت (ت) خالد العامري، التحليل الإحصائي باستخدام برامج SPSS، دار الفاروق للنشر والتوزيع، مصر، 2009، ص 138.

8- Frederick J. Gravetter and Larry B. Wallnau , op.cit. , PP. 626 - 628.

9- J. Cohen , Statistical Power analysis for behavioral Science , SAN Diego , CA. Academic Press , 1988.

10- Frederick J. Gravetter and Larry B. Wallnau , op.cit. , PP. 629 - 633.

11- Ibid , P. 632.

12- Ibid , P. 633.

ثانيا: المصادر:

1- Frederick J. Gravetter and Larry B. Wallnau , Statistics for The Behavioral Sciencies , 8th ed , Wadsworth Cengate Learing , USA , 2010.

2- George Argyrous , Statistics for Social and Health Research , Sage Publications ,London , 2001.

3- J. Cohen , Statistical Power analysis for behavioral Science , SAN Diego , CA. Academic Press , 1988.

4- J. Richard Kendrick , Social Statistics: An Introduction using SPSS , Second ed , USA , 2005.

5- R. Mark Sirikin , Statistics for Social Science , Sage Publications , International Oaks , London , New Delhi , 1995.

6- جولي بالانت (ت) خالد العامري، التحليل الإحصائي باستخدام برامج SPSS، دار الفاروق للنشر والتوزيع، مصر، 2009 م.

7- عبد الله عامر الهمالي، التقنيات الإحصائية ومناهج البحث، منشورات جامعة قاريونس، 2008 م.

الفصل التاسع عشر

اختبار توزيع ثنائي الحد لعينة واحدة

مقدمة:

تناولنا في الفصول السابقة اختبارات Z و t لمتوسط حسابي لعينة واحدة وقد تم تطبيق الإجراءات الإحصائية لهذين الاختبارين على السؤال البحثي المطروح الذي يوجه استقصاءنا لتوزيعات النزعة المركزية والمتغير الذي نرغب في دراسته والذي تم قياسه على المستويين ذي المسافات والنسبي. ويطلق على مثل هذه الاختبارات، الاختبارات المعلمية باعتبارها تقوم باختبار الفروض حول معلمات مجتمع ما (في هذه الحالة المتوسط).

نشير هنا، إلى أن هناك كثيراً من الأحوال التي نرغب فيها في معرفة مظاهر توزيعات متغير ما أكثر من التوزيعات المتعلقة بالمتوسط الحسابي، كالتوزيع التكراري. ويمكننا حساب العديد من الإحصاءات الوصفية لتلخيص بيانات بحثية، هنا يمكننا القول، بأن هذه الإحصاءات الوصفية ذات العلاقة توظف لمساعدة الباحث في الإجابة على الأسئلة البحثية المطروحة.

عود على بدء إذا ما أخذنا على سبيل المثال، القضية التي تم طرحها سلفاً، وهي السياسة التي انتهجتها الهيئة الوطنية للمعلومات حول مسألة التمويل للمنطقة والذي اعتمد فيها المعيار العمري (المتوسط العمري) لمجتمع يتعدى متوسطه 40 سنة. وبوضوح فإن هذه السياسة هي التي توجه تحليلنا لقيمة المتوسط للمتغير الذي نرغب في دراسته وهو – العمر – ولنفترض جدلاً أن الهيئة العامة للصحة قد غيرت من سياستها وبشكل مفاجئ قررت اعتماد مخصصات إضافية لتلك المنطقة فيما يتعلق بالخدمات الصحية فقط إذا كان 20 % أو أكثر من السكان في هذه المنطقة تتجاوز أعمارهم 40 سنة. وفجْأة أصبح المتوسط الحسابي للعمر ليس له علاقة في اعتماد المخصصات، إلا أننا لازلنا نقوم بحساب المتوسط العمري بالرغم من أن المتوسط لن يساعدنا للوصول إلى اتخاذ قرار حول المخصصات.

إن أفضل طريقة لوصف البيانات هو التعامل مع القاعدة السياسية الجديدة للهيئة وذلك بتقسيم العينة إلى أولئك الناس الذين يبلغون 40 سنة من العمر أو أقل، وأولئك الذين تزيد أعمارهم عن 40 سنة، وبعد ذلك نقوم بحساب نسبة كل فئة من هاتين الفئتين، وكنتيجة لذلك يمكننا أن ننظم بياناتنا في نمط بسيط من التوزيعات التكرارية يطلق عليه "اختبار توزيع ثنائي الحد". فالتوزيع ذو الحدين له قيمتين أو فئتين: مثل النوع يمثل توزيع الذكور والإناث[1].

تجدر الإشارة، إلى أن المتغيرات المقاسة على مستويات: المقياس الترتيبي والمقياس ذي المسافات والنسبي، يمكن اختزالها إلى متغيرات تصنيفية أو ثنائية.

البيانات الاسمية:

إن المتغير الاسمي لا يمتلك في جوهره فئتين فقط في هذه الحالة يمكن اختزاله إلى توزيع ذي حدين من خلال تحديد عدد الحالات التي تقع أو لا تقع داخل فئة بعينها أو مجموعة متألفة من الفئات. فعلى سبيل المثال، إن توزيع الحالات المقاسة على المستوى الاسمي طبقاً للانتماء الديني يمكن تحديدها من خلال خمس فئات:

1- الكاثوليك.

2- البروتستانت.

3- اليهود.

4- الأرثوذكس.

5- الإسلام.

وان هذه الفئات يمكن اختزالها إلى توزيع ثنائي بإحدى الطريقتين:

1- الإشارة إلى نسبة الحالات التي تقع أو لا تقع في أحد هاتين الفئتين الموجودة مثل: الكاثوليك وغير الكاثوليك أو مثل المسلمين وغير المسلمين.

2- من خلال توليد فئتين مختلفتين تماماً، وذلك من خلال توحيد هذه الفئات مثل الكاثوليك وغير الكاثوليك، المسلمون وغير المسلمين، انظر الجدولين التاليين:

جدول (19 ـ 1) الانتماء الديني

الديانة	التوزيعات التكرارية
الكاثوليك	20
البروتستانت	15
الأرثوذكس	12
الإسلام	12
اليهودية	7

الديانة	التوزيعات التكرارية
الكاثوليك	20
البروتستانت	15
الأرثوذكس	14
الإسلام	10
اليهودية	7

جدول (19 ـ 2) الانتماء الديني

الديانة	التوزيعات التكرارية	النسبة
الكاثوليك	20	% 30
غير الكاثوليك	46	% 70

الديانة	التوزيعات التكرارية	النسبة
المسلمون	10	% 15
غير المسلمين	56	% 85

المصدر: George Argyrous, Statistics for Social & Health Research, with a Guide to SPSS , SAGE Publications, London, 2001 , P. 304.

البيانات الترتيبية والبيانات ذات المسافات والنسبية[2]:

البيانات الترتيبية أو البيانات ذات المسافات المتساوية والنسبية يمكن اختزالها إلى توزيعات ثنائية (ذات الحدين) وذلك من خلال تحديد عدد الحالات التي تقع فوق أو تحت قيم محددة على المقياس. على سبيل المثال: إذا كانت لدينا قائمة بدرجات امتحان مادة الإحصاء الاجتماعي عندئذٍ بإمكاننا اختزال هذه الدرجات إلى توزيع ذي حدين، وذلك من خلال اختيار نسبة 50 % كخط التقسيم وتنظيم الدرجات إلى ناجح وراسب.

توزيع المعاينة لنسب عينة[3]:

عندما يُتِمُّ الباحث تنظيم بياناته في شكل توزيع ثنائي ويقوم بحساب النسب المتعلقة بالحالات في كل فئة من هاتين الفئتين، عندئذٍ بإمكانه إجراء اختبار الاستدلال لهذه النسب. ولفعل ذلك، ينبغي علينا معرفة توزيعات المعاينة لنسب العينة.

في الفصول السابقة، كان الاهتمام منصباً على إجراء الاستدلال من متوسط العينة على متوسط المجتمع الذي سحبت منه هذه العينة؛ ولفعل هذا الاستدلال فقد تم بناء توزيعات المعاينة لمتوسطات العينة. ويسمح لنا هذا التوزيع تقييم الاحتمالية في الحصول على متوسط العينة الفعلي من مجتمع بقيمة افتراضية محددة (الفرض الصفري).

تجدر الإشارة، عند العمل بالتوزيع الثنائي فإن الإحصاء الوصفي المحسوب من العينة لم يعد متعلقاً بالمتوسط الحسابي بقدر ما يكون متعلقاً بنسبة الحالات التي تقع داخل إحدى الفئتين المحتملة للمتغير. وبعد حساب نسبة العينة حينئذٍ ينبغي علينا بالضرورة القيام بعملية الاستدلال حول النسبة المتعلقة بالمجتمع ككل. وعليه، فالضرورة تقتضي سبر غور خاصيات توزيعات المعاينة لنسب العينة: إن توزيعات نسب المعاينة سوف تظهر من خلال إعادة تكرار العينات العشوائية ذات الأحجام المتساوية. على سبيل المثال، يمكننا معرفة أن 50 % من كل الطلاب في الجامعة (افتراضية) هم من الذكور وأن 50 % من الإناث وبالرغم من هذا، إذا ما تم سحب عينة من 100 طالب من طلاب الجامعة فإنه ليس بالضرورة أن نحصل من خلال هذه العينة على 50 % من الذكور و 50 % من الإناث، فقد يكون التباين العشوائي مسئولاً عن بعض العينات التي تحتوي على عدد

أكبر قليلاً من الإناث، بينما في عينات أخرى سوف تحتوي على عدد أكبر قليلاً من الذكور. لكنه يمكننا القول، بأن معظم هذه العينات المكررة سوف تحتوي على نسبة كل نوع إما مساوٍ أو قريب من 50 %، بمعنى آخر، بينما يوجد بعض التباين في التوزيعات المكررة لنسب العينة، فإن هذه النسب سوف تتجمع حول القيمة الحقيقية للمجتمع وهي 50 %. وإذا سحبنا عدداً لا متناهياً من العينات العشوائية بأحجام متساوية من مجتمع، وقمنا بحساب نسبة الحالات في كل حالة تمتلك قيمة محددة للتوزيع ذي الحدين، فإن توزيع المعاينة لنسب هذه العينات سوف تحتوي على الخاصيات التالية:

1- إن توزيع المعاينة سيكون قريباً من التوزيع المعتدل بنسبة وسيط مساوٍ لقيمة المجتمع. فتوزيع المعاينة يكون فقط، قريباً من المعتدل، لأن التوزيع الثنائي هو متغير منفصل. ولما كان المنحنى الطبيعي هو منحنى متصل، ومع ذلك كلما كان حجم العينة كبيراً، كانت التوزيعات تقارب التوزيع المعتدل.

2- إن الخطأ المعياري لتوزيع المعاينة يمكن تعريفه بالمعادلة التالية:

$$\sigma p = \sqrt{\mathbf{pu}\left(\frac{\mathbf{1 - Pu}}{\mathbf{N}}\right)}$$

حيث إن: Pu تساوي نسبة المجتمع.

تعتبر هذه المعلومات المتعلقة بهاتين النقطتين السابقتين في غاية الأهمية كما تبين لنا في الفصول السابقة.

إن معرفتنا بكل الاحتمالات الممكنة لنسب العينة التي جاءت من مجتمع معين تسمح لنا بحساب الاحتمالية في الحصول على نتيجة عينة محددة من مجتمع بقيمة افتراضية، على سبيل المثال، إذا كانت عينة ما تحتوي على 60 % من الإناث، فإنه باستطاعتنا حساب الاحتمالية بأن هذه النتيجة كانت نتيجة خطأ المعاينة عند عملية سحب العينة من المجتمع الذي تصل فيه نسبة الإناث إلى 50 % فقط. إنه بالتحديد ذلك النوع من السؤال الذي سنتعامل معه، صمم اختبار Z لنسب العينة.

اختبار لنسبة توزيع اختبار ثنائي الحد[4]:

بالرغم من وصفنا للبيانات من خلال تنظيمها وفقاً للتوزيع ثنائي الحد بدلاً من حساب المتوسط الحسابي لهذه البيانات فالإجراءات المتبعة في الاستدلال من العينة على المجتمع هي إجراءات مماثلة. فمن الناحية العملية، فإن الخطوات المتعلقة باختبار الفرض للنسب هي نفسها التي تم اتباعها في اختبار الفرض للمتوسط الحسابي.

لقد أجرينا الاختبار الاستدلالي كما تم إجراؤه في الفصلين المتعلقين باختبار Z واختبار t لعينة واحدة. إلا أننا من خلال هذا الفصل، فإن اختبار الاستدلال سيعتمد على نسبة العينة الواقعة في واحدة من الفئتين من التوزيع الثنائي، بدلاً من الاعتماد على المتوسط الحسابي للعينة. وبما أن توزيعات المعاينة هي توزيعات طبيعية، فإننا سنجري اختبار Z للفرق بين نسب العينة والقيمة المفترضة (هو إجراء مشابه لما قمنا به عند اختبار Z للفرق بين متوسط العينة ومتوسط المجتمع المفترض).

إن المعادلات المحددة لحساب Z للنسب تكون كالتالي:

$$Z = \frac{(Ps - 0.5) - Pu}{\sqrt{\frac{Pu(100 - Pu)}{N}}} \quad \text{العينة}$$

حيث إن: $Ps > Pu$

أو:

$$Z = \frac{(Ps + 0.5) - Pu}{\sqrt{\frac{Pu(100 - Pu)}{N}}} \quad \text{العينة}$$

حيث إن: $Ps < Pu$

Ps نسبة العينة.

Pu نسبة المجتمع.

لاحظ أن إشارة الجمع والطرح لنسبة 0.5 إلى أو من نسبة العينة في هاتين المعادلتين قد حُدِّدَتْ بشكل صارم لأن التوزيع ذا الحدين ليس موزعاً توزيعاً طبيعياً، وبالتالي فإن

عملية الجمع أو الطرح لـِ 0.5 (التصحيح المتواصل) تمدنا بأفضل قيمة تقريبية للتوزيع الطبيعي. ففي العينة التي تزيد عن 30 مفردة فإن هذا التقريب سيكون مناسباً ودقيقاً. أما على الجانب الآخر، إذا كانت العينة أقل من 30 مفردة، فإن التقريب لا يكون دقيقاً، وبالتالي، فإن اختيار احتمالية دقيقة لتوزيع ثنائي يمكن استخدامه.

مثال: نفترض أننا نرغب في معرفة معدل البطالة في منطقة معينة أصابها ركود اقتصادي صعب، مقارنة بباقي أجزاء البلاد. وأن الباحث على معرفة بأن معدل البطالة على مستوى البلاد ككل يصل إلى 11 % وقد قرر سحب عينة من 120 فرداً. وطرح عليهم السؤال التالي: ما إذا كان هؤلاء الأفراد خارج سوق العمل أم لا؟ وقد كانت النتائج المستخلصة من هذه العينة، أن 18 فرداً من إجمالي العينة أقروا بأنهم عاطلون عن العمل. وقد فرغت إجابات هؤلاء المبحوثين على السؤال البحثي المطروح في توزيع ثنائي كما هو موضح في الجدول التالي:

جدول (19 ـ 3): توزيع المبحوثين حسب الوضع المهني

الوضع المهني	التكرار	النسبة المئوية
يعمل	102	% 85.0
لا يعمل (عاطل عن العمل)	18	% 15.0
المجموع	120	% 100.0

المصدر: George Argyrous, Statistics for Social & Health Research, OP. Cit. , P. 307.

هل هذه البيانات حقاً تشير إلى أن هذه المنطقة قد تعرضت لركود اقتصادي صعب؟

الخطوة الأولى: صياغة الفرض الصفري والفرض البديل:

Ho الفرض الصفري: إن نسبة البطالة في هذه المنطقة المحلية مساوية لنسبة البطالة في باقي البلاد.

$$Ho = P_u = \% 11$$

Hi الفرض البديل: إن المنطقة المحلية لديها أعلى نسبة من الأفراد العاطلين عن العمل عن باقي البلاد.

$$Hi = P_u > \% \ 11$$

الخطوة الثانية: اختيار درجة الدلالة:

إن السؤال البحثي المرغوب فيه، هو نسبة الناس في فئة التوزيع الثنائي (ذي الحدين) على سبيل المثال الأفراد العاطلين عن العمل. وعليه فإننا في هذه الحالة سنستخدم اختبار Z أحادي الجانب للنسب.

الخطوة الثالثة: حساب درجة العينة:

إن بيانات هذه العينة المناسبة من التوزيع الثنائي أعلاه تكون:

$$N = 120$$

$$Ps = \frac{18}{120} \times 100 = 15\%$$

بالتعويض نتحصل على:

$$Z_{\text{العينة}} = \frac{(Ps - 0.5) - Pu}{\sqrt{\frac{Pu(100 - Pu)}{N}}} = \frac{(15 - 0.5) - 11}{\sqrt{\frac{11(100 - 11)}{120}}}$$

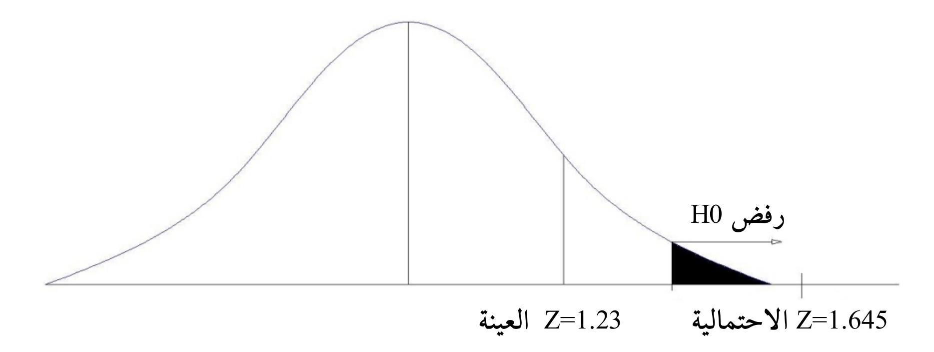
$$= 1.23$$

الخطوة الرابعة: اختيار الدرجات الحرجة والمنطقة الحرجة:

لقد تم اعتماد مستوى الدلالة لهذا المثال بـِ 0.05 (α) ولشك الباحث أن هذه المنطقة المحلية قد تعرضت لركود اقتصادي صعب مما حدا به لاستخدام اختبارٍ أحاديّ الجانبِ. ولما كان اتجاه الفرق قد حدد من خلال الفرض البديل بأن معدل البطالة في هذا المجتمع المحلي يكون أعلى من المعدل الوطني العام. وعليه فإن الجانب المناسب من توزيع المعاينة لهذا الاختبار هو الجانب الأيمن. ومن خلال استخدام الجدول تحت المنحنى الطبيعي المعياري، فإن درجة Z الحرجة هي 1.645 +.

الخطوة الخامسة: اتخاذ القرار:

إن قيمة Z المحسوبة تساوي 1.23 وهي بالتالي أقل من درجة Z الاحتمالية 1.645، انظر الشكل التالي:



شكل (19 ـ 1) درجات العينة والمنطقة الحرجة

بالرغم من أن معدل البطالة في العينة أكثر منه على المستوى الوطني، إلا أن الفرق بين هذين المعدلين ليس كبيراً يجعلنا نقر بأن السكان في هذه المنطقة يكابدون معدلاً عالياً من البطالة. وعليه لا يمكننا رفض الفرض الصفري على مستوى دلالة 0.05.

إجراء اختبار ثنائي الحد Binomial Test باستخدام SPSS:

1- من القائمة الموجودة في الجزء العلوي من الشاشة اختر:

Analyze ⟸ Nonparametric Test ⟸ Binomial

2- من قائمة المتغير انقر على Employment Status .

3- انقر على ◄ يقوم بلصق Employment Status في Test Variable List.

4- في المربع المقابل لِـ Test Proportion تطبع 11.

5- انقر على ok.

فيما يلي المخرجات الناتجة عن هذا الإجراء:

Npar Tests Binomial Test

	Category	N	Observed Prop.	Test Prop.	Asymp. Sig. One taild
Employment Status Group 1	UN Employed	18	.15	.11	.105[a]
Group 2	Employed	102	.85		
TOTAL		120	1.00		

a Based ON Z approximation

المصدر: George Argyrous, Statistics for Social & Health Research, with a Guide to SPSS , SAGE Publications, London , 2001 , p. 309.

شكل رقم (19 ـ 1) مخرجات SPSS لاختبار Z الثنائي

تفسير مخرجات SPSS لاختبارات عينة واحدة لتوزيع ثنائي الحد:

من خلال المربع أعلاه يتبين لنا أن مخرجات SPSS أنتجت لنا الاختبار في شكل تناسب بدلاً من النسب. وهذه عملية لا تؤدي إلى اختلاف جوهري حيث إنه يمكن تحويل التناسب إلى نسب وذلك من خلال تحريك الفاصلة العشرية Decimal Point مكانين إلى اليمين.

في مربع اختبار التوزيع ثنائي الحد Binomial test لدينا عمود يحتوي على تكرار الحالات لكل فئة من فئات التوزيع ثنائي الحد. وكذلك عمود يشير إلى التوزيعات النسبية Relative Frequencies كتناسب Proportions، والعمود الأخير يشير إلى الدلالة المقاربة Asym. Sig.(1-taild) اختبار أحادي الجانب وهو عمود مهم لغرض الاختبار الاستدلالي. وبالرغم من أنه لن يعطي قيمة Z (العينة) فقد أعطيت احتمالية أحادي الجانب المرتبطة به. وهنا احتمالية أحادي الجانب 105. التي تشير إلى أنه إذا كان الفرض الصفري فرضاً صحيحاً على الأقل 1 في 10 عينات ستكون معدلات البطالة 15 % أو أكثر. إن الادعاء أن الفرض الصفري فرضُ صحيحُ لا نستطيع رفضه.

تجدر الإشارة إلى أن مخرجات SPSS دائماً تعطينا اختباراً أحادي الجانب عندما تحدد نسبة الاختبار بدلاً من عدم وجود قيمة 0.05. وإذا كان الفرض البديل يتطلب

اختباراً ثنائي الجانب at two-tail test عندئذ وبكل بساطة تضاعف احتمالية اختبار أحادي الجانب. وعلى سبيل المثال إذا كان توزيع المعاينة لاختبار أحادي الجانب لدرجة Z تحتوي على 0.105 من المسافة تحت المنحنى الطبيعي، فإن اختبار ثنائي الجانب يصل إلى 0.21 من المسافة تحت المنحنى الطبيعي(5).

تقدير نسبة مجتمع(6):

تقدير نسبة مجتمع تعتبر عملية شائعة الاستخدام في مسوح الرأي العام وعادة ما نقرأ في أحوال كثيرة في الصحف أن نسبة معينة من المجتمع تفضل هذه القضية أو تلك، وعادةً لا تمثل هذه النسب كل أفراد المجتمع ولكن من خلال عينة يتم سحبها لهذا الفرض، الأمر الذي يتطلب منا تقدير قيمة المجتمع من نتائج العينة المسحوبة من هذا المجتمع وتستخدم المعادلة التالية لبناء فترة الثقة للنسب.

$$Ci = Ps \pm Z\sqrt{\frac{Ps(100 - Ps)}{N}}$$

إذاً يمكن استخدام هذه المعادلة لبناء فترات الثقة لتقدير نسبة الناس الذين يُدْلُونَ – على سبيل المثال – بأصواتهم لأحد الأحزاب المتصارعة على السلطة، واستناداً على بيانات العينة التالية يمكننا بناء فترات الثقة هذه.

$$N = 120$$

$$Ps = \%\ 15$$

$$\alpha = 0.05\ ,\ Z = (1.96)$$

من هذه البيانات يمكننا حساب فترة الثقة:

$$Ci = Ps \pm Z\sqrt{\frac{Ps(100 - Ps)}{N}} = 15 \pm 1.96\sqrt{\frac{15(100 - 15)}{120}}$$

$$= 15 \pm 6.4$$

وهكذا يكون الحد الأدنى لفترة الثقة 8.6 في المائة (6.4 – 15) والحد الأعلى لفترة الثقة 21.4 في المائة (6.4 + 15).

مثال إضافي لزيادة التوضيح[7]:

عينة من 500 طالب كانوا قد سجلوا في مسار تحليل البيانات بالدراسات العليا مقابل رسوم دراسية. منهم 55 % ينحدرون من أسر ذات خلفية اجتماعية – اقتصادية متدنية. ما هي نسب التقدير لكل طلاب الدراسات العليا الذين قاموا بدفع الرسوم الدراسية مقابل هذا المسار، وينحدرون من أسر ذات وضع اقتصادي واجتماعي متدنٍّ؟

$$n = 500$$

$$Ps = \frac{55}{500} \times 100 = \%11$$

$$\alpha = 0.05\,,\ Z = (1.96)$$

$$Ci = Ps \pm Z\sqrt{\frac{Ps(100 - Ps)}{N}}$$

$$= 11 \pm 1.96\sqrt{\frac{11(100 - 11)}{500}}$$

$$= 11 \pm 2.7$$

بمعنى آخر، إن نسبة الطلاب ذوي الخلفية الاقتصادية والاجتماعية المتدنية في السكان تقع بين (8.3 % و 13.7 %) بمستوى ثقة (95 %).

الاستدلال باستخدام فترة الثقة للنسب[8]:

إن السبب وراء إجراء المسح السابق هو معرفة ما إذا كان اعتماد الرسوم الدراسية لمسارات برنامج الدراسات العليا يؤدي إلى نتيجة معاكسة على المجموعات المحرومة من الطلاب.

إن الإحصاءات السكانية المتوفرة لطلاب الدراسات العليا قبل إدخال اعتماد الرسوم الدراسية الجامعية، كانت تشير إلى 16.8 % من طلاب الدراسات العليا ينحدرون من أسر ذات وضع اقتصادي واجتماعي متدنٍّ. هل يمكننا القول، بأن إدخال الرسوم الدراسية قد أثرت على دخول الطلاب الفقراءً للجامعة، نريد أن نقارن عدد طلاب الدراسات العليا قبل إدخال الرسوم الدراسية بعدد الطلاب بعد الإدخال. ونود

أن نشير إلى أنه لدينا معلومات قبلية فيما يتعلق بالرسوم الطلابية ولكن لدينا فقط تقديرات بعدية للرسوم المتعلقة بطلاب الدراسات العليا استناداً على عينة، وعلى أية حال يمكننا أن ندرك تقدير نسبة المجتمع للرسوم البعدية والتي لا تتضمن 16.8 %. إن نسبة الطلاب المحرومين فيما يتعلق بالرسوم القبلية لا تقع ضمن 95 % فترة ثقة ضمن ما قدرناه. حيث تكمن في نسبة الطلاب بعد إكمال الرسوم الدراسية البعدية لدى الطلاب الذين جاءوا من خلفيات اجتماعية واقتصادية متدنية، بمعنى آخر، استناداً على نتائج العينة يمكننا القول بأن نسبة الطلاب المحرومين في فترة قبل اعتماد الرسوم الدراسية كانت نسبة عالية وذات دلالة أكثر منها بعد إدخال الرسوم الدراسية. وعليه، يمكننا القول، بأن إدخال الرسوم الدراسية كان له تأثير على المجموعة المحرومة من الطلاب.

عليه، يمكننا إجراء اختبار الاستدلال لمعرفة ما إذا كنا قادرين على رفض الفرض الصفري حول نسبة مجتمع من خلال بناء فترة ثقة ويعتبر هذا بديلاً لاختبار الفرض الذي تعلمناه من خلال الفصول السابقة. في الحقيقة سوف نقوم بإجراء اختبار Z لعينة واحدة للنسب، وللوقوف على هذه النقطة بشكل أكبر، لنبيّن أن النتيجة ستكون واحدة عند استخدام فترة الثقة للوصول إلى قرار حول الفرض:

العينة

$$Z = \frac{(Ps - 05) - Pu}{\sqrt{\frac{Pu(100 - Pu)}{N}}}$$

$$= \frac{(11 - 05) - 16.8}{\sqrt{\frac{(16.8)(100 - 16.8)}{500}}}$$

$$= 3.8$$

في ضوء اختبار ثنائي الجانب، فإن قيمة Z الحرجة بمستوى دلالة 0.05 (مساوية لـ 95 % مستوى ثقة) ستكون ±1.96 وبوضوح، فإنه بإمكاننا أن نرفض الفرض الذي مفاده أن هذه العينة جاءت من مجتمع يمثل نسبة المحرومين فيه من الطلاب 16.8 %.

العلاقة بين مربع كاي واختبار Z ثنائي الحد:

ناقشنا في الفصل السابق بشكل واسع اختبار مربع كاي للاستقلال، باعتباره من التقنيات الإحصائية لاختبار الدلالة الأكثر شيوعاً واستخداماً في مجال البحث الاجتماعي، ويعزى سبب شيوع استخدام مربع كاي أنه يمكن تطبيقه في مواقف يكون فيها لدينا بيانات مصنفة تصنيفاً اسمياً وترتيبياً، ناهيك عن رغبتنا في الحصول على التوزيع التكراري عبر فئات المتغير الخاضع للدراسة. ويعتبر هذا الموقف موقفاً شائعاً في البحث الاجتماعي.

إن اختبار مربع كاي يفحص توزيع الإجابات في جدول ثنائي يقيم ما إذا كان نمط التبعية موجوداً في حالة الجدول الثنائي 2 × 2 (عندما يكون كلا المتغيرين ثنائي الحد binominal). فإن اختبار Z للنسب يمكن إجراؤه على البيانات نفسها؛ وفي الحقيقة أن كلا الاختبارين متكافئُ في طرق تحليل نفس البيانات، والوصول إلى النتيجة نفسها أي ($X^2 = Z^2$).

حقاً إن اختبار Z للنسب يمكن اعتباره حالة خاصة لاختبار مربع كاي. ولما كان اختبار مربع كاي اختباراً شائع الاستخدام في البحوث، أصبح من الأهمية بمكان معرفة آليات وطرق حسابه.

في هذه الجزئية من هذا الفصل سوف نتعامل مع مثال لاختبار Z لنسب عينة، وبعد ذلك نجري اختبار مربع كاي لنبرهن على أن النتائج التي نتوصل إليها ستكون واحدة وأن أي منهما يمكن استخدامه.

مثال تطبيقي:

لقد تم إجراء مسح (افتراضي) لاستقصاء مستوى الدعم لإصلاح الرعاية الاجتماعية، وما إذا كان هناك تفاوت Varies حسب العمر في هذا المسح. لقد تم ترتيب المبحوثين طبقاً للفئة العمرية: تحت 45 سنة أو 45 سنة أو أكثر. وقد تم طرح سؤال على كل مبحوث مفاده ما إذا كان ينبغي على الدولة أن تقوم بجهد أكبر للتخفيف من الفقر. وقد تم وضع هذا السؤال بصيغة مبسطة (نعم - لا). وقد تمت صياغة الفرض الصفري بأن

نسبة الذين أجابوا بنعم تحت 45 سنة (P_1) تكون مساوية لنسبة أولئك الذين أجابوا بنعم 45 سنة أو أكثر (P_2):

$$Ho: P_1 = P_2$$

إذا كانت هذه الفرضية صحيحة، فإن العينات المسحوبة من مثل هذه المجتمعات ففي العادة سوف تعكس صورة متساوية. بمعنى آخر، أن الفرق بين أي نسب للعينتين، إذا لم يكن هناك فرق بين المجتمعات ينبغي أن يكون هذا الفرق صفراً أو قريباً منه.

تجدر الإشارة إلى أن هذا الأمر لن يكون دائماً على هذا النحو. فالعينات ليست دائماً تعكس خصائص المجتمع التي سحبت منها. فالتباين العشوائي قد يقودنا إلى أن نختار زيادة إضافية طفيفة من صغار السن الذين يؤيدون إصلاح نظام الرعاية الاجتماعية، وكذلك نختار مجموعة إضافية طفيفة من كبار السن الذين يعارضون هذا الإصلاح.

إن مثل هذه الاختبارات الإضافية الطفيفة تسبب إلى حد كبير اختلافاً في نسب العينة. وهذا يعني أنه إذا كان هناك فرق بين نسب العينتين، فإننا لا نستطيع بشكل تلقائي أن نقرر أن هذا الفرق يعكس ما يبطن من فرق في المجتمعات. وعليه فإن الفروق الكبيرة بين نسب العينات يكون أقل احتمالاً بأن تعزو هذا الفرق إلى الفرضية العشوائية. إن اختبار Z للنسب يعطي احتمالية دقيقة أن مثل هذه الوقائع احتمالية حدوثها تكون بعيدة.

إن المسح الذي يحتوي على 600 شخص تحت الفئة العمرية 45 سنة و 400 شخص أعمارهم 45 سنة أو أكثر. إن النسبة لكل مجموعة أجابت "بنعم" (أي يتوجب على الدولة بذل جهد كبير للتخفيف من الفقر) جاءت كالتالي:

1- تحت سن 45 سنة : $P_1 = \frac{490}{600} \times 100 = 82\%$

$$N_1 = 600$$

2- 45 : 45 سنة أو أكبر : $P_2 = \frac{232}{400} \times 100 = 58\%$

$$N_2 = 400$$

السؤال الذي ينبغي طرحه في هذا السياق هو هل هذه النسب تعكس ما يتضمنه الفرق بين المجموعات العمرية في هذه القضية المطروحة؟ لتحديد ذلك نبدأ بالمعادلة التالية:

$$Pu = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{600(82) + 400(58)}{600 + 400}$$
$$= 72.4\%$$

إن هذه النتيجة التي توصلنا إليها تسمح لنا بأن نحدد الخطأ المعياري Standard error لتوزيع المعاينة لكل فروق العينة المحتملة. وإن خطأ معيارياً واحداً يمكن تعريفه بـِ:

$$\sigma P - P = \sqrt{Pu(100 - Pu)}\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$
$$= \sqrt{72.2(100 - 72.2)}\sqrt{\frac{600 + 400}{600(400)}}$$
$$= 2.9$$

إن الفرق الحقيقي بين هاتين العينتين فيما يتعلق بدرجة Z هو:

$$Z = \frac{p_1 - p_2}{\sigma P - P} = \frac{82 - 58}{2.9} = 8.3$$

إن درجة Z هذه تعتبر ذات دلالة على مستوى ألفا $\alpha = 0.01$ وعليه نرفض الفرض الصفري H0 الذي مفاده لا فرق $P_1 = P_2$ وأنه يمكننا المجادلة بأن دعم الدولة لمساعدة الفقراء يتباين وفقاً للعمر [9].

إن الطريقة البديلة لتحليل هذه البيانات هو أن يتم تنظيمها في جدول ثنائي 2×2 كما هو موضح أدناه:

جدول (19 ـ 4): الاتجاه نحو سياسة الدولة نحو إصلاح نظام الرعاية الاجتماعية حسب مجموعة العمر

الاتجاه	مجموعة العمر		
	تحت 45 سنة	45 سنة أو أكثر	المجموع
لا	(166.8) 110	(111.2) 168	(27.8%) 278
نعم	(433.2) 490	(288.8) 232	(72.4%) 722
المجموع	600	400	1000

المصدر: George Argyrous, Statistics for Social & Health Research,op.cit, , p.417.

تشير الأرقام بين الأقواس إلى القيم المتوقعة وذلك استناداً إلى النسبة الكلية للمبحوثين الذين أجابوا "بنعم" أو "لا"، مع ملاحظة أن نسبة 72.4 في المائة من كل المبحوثين قد وافقوا على ضرورة إصلاح نظام الرعاية الاجتماعية. ومن خلال هذا العدد يمكننا حساب عدد المبحوثين تحت 45 سنة وكذلك حساب عدد المبحوثين ذوي الأعمار 45 سنة أو أكثر الذين يتوقع موافقتهم على إصلاح نظام الرعاية الاجتماعية. إن نسبة 72.4 في المائة هي النسبة ذاتها التي وردت في اختبار Z للنسب لعينتين كنقطة مرجعية لحساب الانحراف المعياري لتوزيع المعاينة. في الحقيقة، أنه عندما نقوم بحساب مربع كاي:

$$X^2 = \sum \frac{(Fo - Fe)^2}{Fe}$$

$$= \frac{(110-168.8)^2}{168.8} + \frac{(168-111.2)^2}{111.2} + \frac{(490-433.2)^2}{433.2} + \frac{(323-288.8)^2}{288.8}$$

$$X^2 = 68 = Z^2 = 8.3$$

من الجدول لتوزيع مربع كاي، فإن الاحتمالية للحصول على هذه القيمة (الأكبر) من مجتمعات مماثلة هي 0.005 – وهي مساوية لتلك الاحتمالية المتعلقة باختبار Z.

إن النتيجة التي تستنتج من هذا هي أنه بينما تكون اختبارات Z لعينتين ثنائية الحد، اختبارات شائعة جداً، وعليه، أصبح الإلمام والمعرفة بهذا الاختبار مهماً جداً، وفي الحقيقة

فإن هذه الاختبارات تمثل حالة خاصة لمربع كاي. وبما أن معادلة اختبار Z هي أكثر إرهاقاً وأن منطقها غير واضح إدراكياً، وبالتالي فإن استخدام اختبار مربع كاي على الأرجح سيكون مناسباً في معظم المواقف[10].

أسئلة للمراجعة:

1- من مجموع البيانات التالية أجرِ اختبار Z للنسب مستخدماً كل من: اختبار أحادي الجانب واختبار ثنائي الجانب بألفا (α) 0.05.

A	B
$P_u = 52$	$P_u = 42$
$P_s = 61$	$P_s = 39$
$n = 110$	$n = 110$

2- عينة من 900 سجين تم مسحها لضمان نجاح برنامج إعادة التأهيل بالسجن. 350 من هؤلاء السجناء أوضحوا بأن البرنامج كان فعالاً للتقليل من احتمالية إعادة ارتكاب الجريمة. لقد كان الهدف من هذا البرنامج هو تحقيق 40 % من معدل النجاح في التقليل من احتمالية العود إلى ارتكاب الجريمة:

أ- باستخدام اختبار Z للنسب، هل يمكننا القول بأن هذا البرنامج قد نجح.

ب- بناء 95 % فترة ثقة لتقدير قيمة المجتمع. كيف يمكن أن تؤكد فترة الثقة نتيجة اختبار Z.

3- أجريت دراسة على 500 شخص. 56 % من هؤلاء يؤيدون تولي المرأة مناصب عليا في الإدارة. ما هي فترة الثقة لـِ 95 % لنسبة كل الناس الذين يؤيدون تولي المرأة مناصب عليا في الإدارة؟ هل يمكنك القول بأن أغلبية الناس يؤيدون هذا الاتجاه؟

4- عينة عشوائية لعدد 60 مصنعاً. وجد منها 15 مصنعاً لا تتقيد بمعايير السلامة البيئية. ما هي فترات الثقة لـِ:

a) 90 %.

b) 95 %.

الهوامش والمصادر:

أولا الهوامش:

1- George Argyrous, Statistics for Social & Health Research, with a Guide to SPSS , SAGE Publications, London, 2001, P. 303.

2- Ibid. , P. 304.

3- Ibid. , PP. 305 - 306.

4- Ibid. , P. 308.

5- Ibid. , P. 310.

6- Ibid. , PP. 311 - 312.

7- Ibid. , P. 312.

8- Ibid. , PP. 312 - 318.

9- Ibid. , P. 417.

10- Ibid. , P. 418.

ثانيا: المصادر:

1- George Argyrous, Statistics for Social & Health Research, with a Guide to SPSS , SAGE Publications, London, 2001.

2- Frederick J. Gravetter and Larry B. Wallnau , Statistics for The Behavioral Sciencies , Wadsworth Cengage Learning , USA , 2010.

3- Joseph F. Healey , The Essentials of Statistics: A Tool for Social Research , Wadsworth Cengage Learning , USA , 2010.

الفصل العشرون

الأساليب الإحصائية للبيانات الترتيبية

اختبار الفرض:

في هذا الفصل سوف نتناول الإجراءات المتعلقة باختبار الفرض من البيانات مقاسة علي المستوى الترتيبي. وأن كل واحد من هذه الاختبارات يمكن النظر إليه كبديل للاختبارات المعلمية التي تناولناه في الفصول السابقة، وأن الاختبارات الأربعة التي سنتطرق لها، والمواقف التي تستخدم فيها هي كالتالي:

1- اختبار مان- وتني Mann-Whitney

يستخدم هذا الاختبار بيانات من عينتين منفصلتين وذلك لتقييم الفرق بين مجتمعَيْن. ومن هنا يمكن اعتبار اختبار مان- وتني اختباراً بديلاً لاختبار t لعينتين مستقلتين.

2- اختبار ولكوكسن Wilcoxon Test

لقد صُمِّمَ هذا المقياس لتقييم الفرق بين معالجتين. وذلك باستخدام بيانات تم الحصول عليها من تصميم متكرر القياسات a repeated-measures، ويعتبر هذا الاختبار اختباراً بديلاً لاختبار t لمقاييس متكررة a repeated t-test.

3- اختبار كروسكال-ويلز Kruskal-Wallis Test

يُسْتَخْدِمْ هذا الاختبار بياناتٍ لتقييم الفروق بين ثلاث عينات أو أكثر مستخدماً بيانات تم الحصول عليها من خلال تصميم المقاييس المستقلة. ويعتبر هذا المقياس مقياساً بديلاً لتحليل التباين أحادي الجانب Anova.

4- اختبار فريدمان The Friedman test

يستخدم هذا الاختبار بيانات تم الحصول عليها من تصميم متكرر القياسات a repeated-measurea design، وذلك لمقارنة الفروق بين ثلاثة أو أربعة أحوال من المعالجة. ويعتبر هذا الاختبار اختباراً بديلاً لتحليل أنوفا ANOVA لمقاييس متكررة.

وتجدر الإشارة إلى أن هذه الاختبارات تكون ملائمة للاستخدام عندما لا يكون في مقدور الباحث أن يفي بمتطلبات الاختبارات البارامترية (المعلمية).

وبشكل عام، إذا كانت البيانات المتحصل عليها بيانات ملائمة لإجراء اختبار Anova أو أي من اختبارات t. حينئذٍ يكون الاختبار المعياري standard test مفضلاً للبيانات الترتيبية البديلة [1].

5- الحصول علي المقاييس الترتيبية:

للحصول علي الرتب من الملاحظة المباشرة، يمكن للباحث أن يبدأ بمجموعة من المقاييس العددية وتحويل هذه الدرجات إلى رتب Ranks. فعلى سبيل المثال، إذا كان لدى الباحث الأطوال الحقيقية لمجموعة من الأفراد، فإنه بإمكانه أن يرتب الأعداد من الأكبر إلي الأصغر. ويسمح هذا الإجراء بتحويل البيانات من بيانات مقاسة على المستوى ذي المسافات والنسبي إلى مقاييس ترتيبية. إن القائمة التالية سوف تلقي الضوء على المميزات وراء استخدام الرتب بدلاً من الدرجات.

1- إن عملية الرتب هي عملية بسيطة. فإذا سألك أحد عن كم طول ابنك، فإنك ببساطة ستجيبه بقيمة عددية محددة مثل: 170 سم، أو 164سم..... إلخ. أو أنك قد تجيبه بأنه أقصر منك بقليل أو يزيد طوله قليلاً عن طولك.

2- إن الدرجات الأصلية يمكن أن تنتهك بعض الافتراضات الأساسية التي تتضمنها بعض الإجراءات الإحصائية. على سبيل المثال، إن اختبارات t واختبارات Anova تفترض أن البيانات قد جاءت من توزيعات طبيعية Normal distributions، أيضاً فإن اختبارات المقاييس المستقلة Independent measures، تفترض أن مجتمعات مختلفة لديها نفس التباين (فرضية تجانس التباين). أما إذا شك الباحث في أن البيانات لا تفي بهذه الافتراضات، عندئذٍ ينبغي عليه تحويل هذه الدرجات إلى رتب، واستخدام التقنيات الإحصائية المصممة للرتب.

3- إن الدرجات الأصلية قد يكون لديها تباين عالٍ؛ ويعتبر هذا التباين المكون الأساسي للخطأ المعياري في المقام المتعلق لمعادلة إحصاء t. ومصطلح الخطأ في مقام معادلة اختبار f. وعليه، فإن التباين الكبير يمكن من خلاله بشكل واضح تقليل احتمالية أن هذه الاختبارات البارامترية (المعلمية) سوف تكتشف لنا فروقاً دالة.

إن تحويل الدرجات إلى رتب، سيقلل بشكل جوهري من عملية التباين، على سبيل المثال، إذا تم ترتيب عشر درجات (10) من 1 إلى 10، فإن الأمر لا يهم كم كانت الدرجات الأصلية للمتغير.

6- ترتيب الدرجات المتعادلة Ranking Tied Scores:

كلما قام الباحث بتحويل الدرجات العددية إلى رتب، فإنه بالتالي قد يجد رقمين أو أكثر يحملان تماماً نفس القيمة. ولأن الدرجات كانت متعادلة، فإن عملية إجراء التحويل يجب أن تولد رتباً تكون أيضاً متعادلة.

إن عملية إجراء تحويل الدرجات المتعادلة إلى رتب، قد تمت مناقشتها عند الحديث عن معامل سبيرمان للرتب. ولكننا قد نعيد هذا الإجراء بشكل مختصر في هذا الفصل. أولاً ينبغي على الباحث أن يقوم بتنظيم الدرجات في قائمة تشتمل على القيم المتعادلة. وثانياً يحدد لكل حالة في القائمة رتبة معينة (الأول، الثاني..... إلخ). ثالثاً وأخيراً، إذا كانت هناك درجتان متعادلتان، فإنه ينبغي علي الباحث أن يقوم بحساب متوسط الرتب المتعادلة، واعتماد قيمة المتوسط كرتبة نهائية. ومجموع الدرجات التالية تبين هذا الإجراء لعدد ثمان درجات $N = 8$:

الدرجات الأصلية	3	4	4	7	9	9	9	12
الوضع الترتيبي	1	2	3	4	5	6	7	8
الرتب النهائية	1	2,5	2,5	4	6	6	6	8

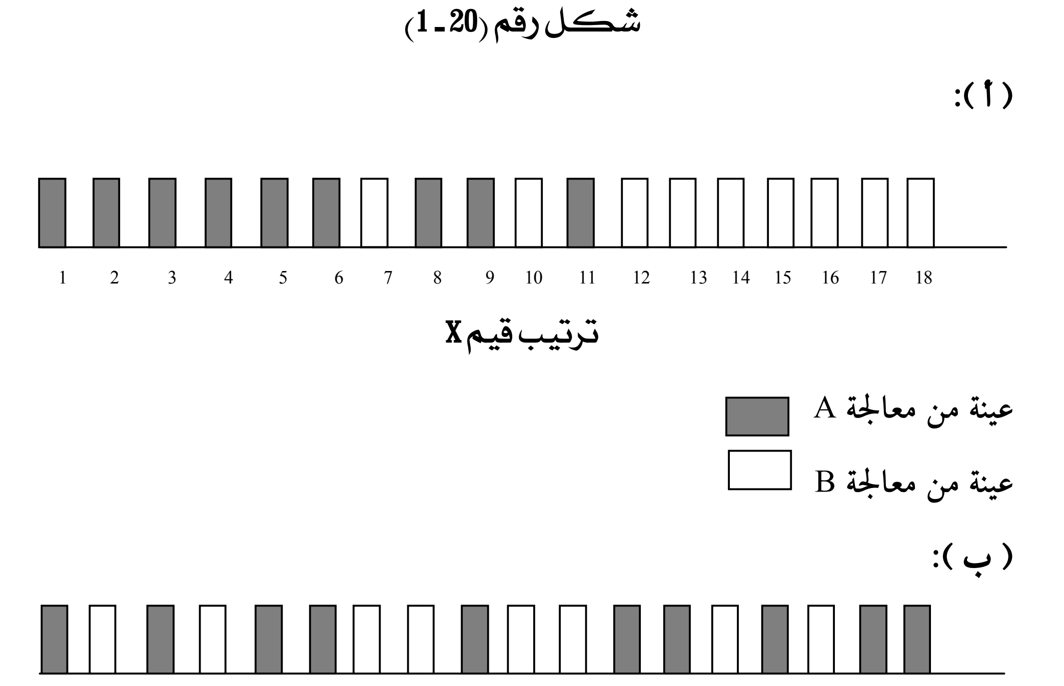
علي السبيل المثال، تحتوي الدرجات الأصلية علي شخصين درجاتهما متعادلة 4 = x وقد أعطي لهاتين الدرجتين رتب 2، 3، وبعد ذلك فإن كلا الشخصين قد أسند إليهما رتبة نهائية مساوية لمتوسط هذه الرتب (2.5=2 / 5=2+3). وهكذا في باقي الدرجات المتعادلة.

أولاً: اختبار مان- وتني The Mann-Whitney U كبديل لاختبار مقاييس t :

لقد صُمِّمَ هذا الاختبار لاستخدام بيانات من عينتين منفصلتين وذلك لتقييم الفرق بين مجتمعين لديها بيانات عددية، وغالباً ما يستخدم هذا الاختبار كبديل للاختبار t. ويتطلب حساب هذا الاختبار أن الدرجات الفردية في العينتين ينبغي أن ترتب، وأن العملية الرياضية لاختبار مان- وتني U تستند علي الملاحظة البسيطة التالية:

إن الفرق الحقيقي بين المجتمعين يجب أن يكون سبباً في وجود الدرجات في واحد من العينات تكون بشكل عام أكبر من الدرجات في العينة الأخرى. وإذا ما تم دمج عينتين لتصبحا عينة واحدة، ويتم ترتيب كل هذه الدرجات في صف واحد. وعندئذٍ فإن الدرجات من عينة واحدة يجب أن تتركز علي الطرف الأخر من الصف. وعلى الجانب الآخر، فإذا لم يكن هناك فرق، فإن الدرجات الكبيرة والصغيرة سوف تدمج بشكل متساوٍ في العينتين، لأنه لا يوجد أي سبب لأي واحد من مجموع الدرجات لأن يكون بشكلٍ منظم أكبر أو أصغر من مجموع الدرجات الأخرى. إن هذه الملاحظة يمكن التدليل عليها من خلال الشكل التالي:

شكل رقم (20 ـ 1)



المصدر: frederick J. GRavetter and larry B. wallnau , statistics for the Behavioral sciences , .8et ed, Wadsworth cengage learning , USA , 2010 , p.688

الفرضية الصفرية المرتبطة باختبار مان-وتني U :

لما كان اختبار مان- وتني U يقارن بين توزيعين (بدلاً من متوسطين)، فإن الفروض المتعلقة بهذا الاختبار تفضي إلى بعض الغموض، ويصاغ الفرض الصفري في إطار فرق متين ومنظم بين المجموعتين اللتَيْن يسعى الباحث للمقارنة بينهما.

الفرض الصفري H_O: لا يوجد فرق بين المعالجتين وعليه، لا يوجد أي غرض من وراء رتب حالة معالجة بأن تكون بشكل منظم أعلي (أو أقل) من رتب حالة المعالجة الأخرى.

الفرض البديل H_i: يوجد فرق بين المعالجتين. وعليه فإن رتب حالة إحدى المعالجات تكون بشكل منظم أعلى (أو أقل) من رتب حالة المعالجة الأخرى.

حساب قيمة اختبارات U:

إن الخطوة الأولى لحساب اختبار مان وتني U:

1- الحصول على عينتين منفصلتين. ونشير إلى عدد الأفراد في العينة الأولي بالرمز N_1،وعدد الأفراد في العينة التالية بالرمز N_2.

2- يتم دمج العينتين معاً N_1+N_2 وبعد ذلك يتم ترتيبهما.

مثال: إذا كان لدينا عينتان منفصلتان $N_1=6$ و $N_2=6$

N_1 = 15، 6، 48، 9، 2، 27

N_2 = 8، 94، 68، 18، 63، 71

أولا: نقوم باستخراج قيمة Uبحساب عدد الدرجات في العينة الأولى التي تسبق الدرجات في العينة الثانية، أي كم درجة في العينة الأولى يكون ترتيبها أقل من أي درجة من درجات العينة الثانية. ومن خلال ترتيب الدرجات تَبَيَّنَ لنا أن الدرجة الأولى في المجموعة الأولى لا يسبقها أية درجة من المجموعة الثانية أي أصغر وهكذا يمكننا الحصول على قيمة U_1 .

$$U_1 = 0 + 0 + 1 + 1 + 2 + 2$$
$$= 6$$

وبنفس الطريقة يمكننا حساب درجات العينة الثانية فنحصل كل قيمة U_2.

$$U_2 = 2 + 4 + 6 + 6 + 6 + 6$$
$$= 30$$

وبذلك يصبح لدينا قيمتان لـِ: U_1 و U_2

المعادلة التالية لهذه البيانات $U_1 + U_2 = n_1n_2$

$$6 + 30 = (6)(6)$$

ويمكننا أيضاً الحصول على قيمة U من نفس البيانات السابقة بإتباع الأتي:

نقوم أولاً بتنظيم هذه الدرجات الخاصة بهاتين العينتين وترتيبهما من أصغر درجة إلى أكبر درجة، ويتم ذلك بعد دمج درجات العينتين معاً.

أما حساب U من العينات الكبيرة، فإنه لما كانت عملية عد النقاط لكي نحدد قيمة مان – وتني هي عملية مضنية لا سيما عندما يكون حجم العينة كبيراً، فإننا نورد المعادلة التالية لتوليد قيمة U لكل عينة، ولاستخدام هذه المعادلة يتطلب الأمر دمج العينتين معاً، ومن ثم ترتيب كل الدرجات. وبعد ذلك ينبغي الحصول علي $\sum R_1$ أي مجموع ترتيب الأفراد في العينة N_1 ويقابلها $\sum R_2$ للعينة N_2 . وعندئذٍ يمكننا حساب قيمة U وفقاً للمعادلة التالية لكلتا العينتين.

$$U_1 = n_1(n_2) + \frac{n_1(n_1+1)}{2} - \sum R_1$$

$$U_2 = n_1(n_2) + \frac{n_2(n_2+1)}{2} - \sum R_2$$

من خلال البيانات في المثال السابق:

الدرجات	2	6	8	9	15	18	27	48	63	68	71	94
العينة	1	1	2	1	1	2	1	1	2	2	2	2
الترتيب	1	2	3	4	5	6	7	8	9	10	11	12

إن مجموع ترتيب الأفراد في العينة N_1 يكون:

$$\sum R_1 = 1 + 2 + 4 + 5 + 7 + 8 = 27$$

ومجموع ترتيب الأفراد في العينة N_2:

$$\sum R_2 = 3 + 6 + 9 + 10 + 11 + 12 = 51$$

وباستخدام المعادلة الخاصة بالعينة N_1 :

$$U_1 = n_1(n_2) + \frac{n_1(n_1 +_1)}{2} - \sum R_1$$

$$= 6(6) + \frac{6(7)}{2} - 27$$

$$= 36 + 21 - 27$$

$$= 30$$

وللعينة الثانية N_2:

$$U_2 = n_1(n_2) + \frac{n_2(n_2 +_1)}{2} - \sum R_2$$
$$= 6(6) + \frac{6(7)}{2} - 51$$
$$= 36 + 21 - 51$$
$$= 6$$

لاحظ أن هذه القيم هي نفس القيم التي تحصلنا عليها باستخدام الطريقة الأولى.

الاختبارات الصفرية من خلال اختبار مان ـ وتني U:

تجدر الإشارة إلى أنه حتى الآن قد تم تطوير طريقة لبيان نظام الترتيب من خلال قيمة عددية. وعليه فإن المشكلة اللاحقة هي أن يقرر الباحث ما إذا كانت قيمة U تقدم لنا البرهان إن كان هناك فرق حقيقي بين المجموعتين. ولما كان الفرض الصفري لاختبار مان - وتني U يصاغ بأنه لا يوجد فرق نظامي بين العينتين اللتين تخضعان للمقارنة؛ أي لا يوجد فرق حقيقي بين المجتمعين اللَّذَيْن تم سحب العينة منهما. في الحالة الراهنة، فإن النتيجة الأكثر ترجيحاً أن العينتين تكونان متماثلتين، وأن قيمة U تكون كبيرة نسبياً. وعلى الجانب الآخر، أنه عندما تصل قيمة U قريبة من الصفر، يكون هناك نزعة نحو رفض الفرض الصفري.

إن التوزيع الاحتمالي لكل قيم U قد تم بناؤه. وأن القيم الحرجة لمستوى الدلالة $\alpha = .05$، ومستوى الدلالة $\alpha = .05$ يمكن الرجوع إليهما من خلال الجدول (7) انظر ملحق (1). وعندما نتحصل على قيمة U من بيانات العينة أصغر أو مساوية للقيمة الجدولية حينئذٍ يمكننا رفض الفرض الصفري (هذا عكس ما تعودنا عليه في الفرض الصفري عند التعامل مع الاختبارات)، فعلى سبيل المثال، في المثال الذي بين أيدينا نجد أن كلتا العينتين لديها: N تساوي 6 درجات، وأن القيمة الجدولية لـِ U تساوي 5 (اختبار ثنائي الجانب) بمستوي دلالة $\alpha = .05$. هذا يعني أن البيانات التي تحصلنا عليها لا تقدم لنا البرهان الكافي لكي نصل إلى نتيجة مفادها أن هناك فرقاً دالاً بين المجموعتين.

التقدير التقريبي الطبيعي لاختبار مان ـ وتني U:

عندما تكون لدينا عينتان كبيرتان N=20، ويكون الفرض الصفري فرضاً صحيحاً، فإن توزيع إحصاء U يميل إلى أن يكون التوزيع قريباً من الشكل الطبيعي approximate normal shape. في هذه الحالة فإنه باستطاعتنا تقييم الفروض المرتبطة باختبار مان- وتني من خلال استخدام إحصاء درجة Z-Score ، أي تحويل درجة U إلى درجة معيارية (Z)، ويمكننا هنا الإشارة إلى أن جدول القيم الحرجة لاختبار مان - وتني لا يتضمن قيماً لعينات أكبر من N=20، ويرجع السبب في ذلك إلى أن التقدير التقريبي الطبيعي نموذجيٌّ Typically يستخدم مع العينات الكبيرة.

إن الإجراء المتبع لهذا التقدير التقريبي الطبيعي هو كالتالي:

1- إيجاد قيم U لعينة N_1، والقيمة N_2 كما فعلنا سلفاً، وتعتبر قيمة U هي القيمة الأصغر من هاتين القيمتين.

2- عندما يكون حجم العينتين كبيراً نسبياً (عدد الأفراد 20 أو أكثر) فإن توزيع إحصائي اختبار مان - وتني U يميل ليشكل توزيعاً طبيعياً:

$$U = \frac{n_1 n_2}{2} \quad \text{و} \quad \sigma = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

ويمكننا الحصول علي U من بيانات العينة التي يمكن تحديدها في هذا التوزيع مستخدمين درجة Z (Z-Score) :

$$Z = \frac{x - u}{\sigma} = \frac{\frac{U - n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

3- وباستخدام وحدة الجدول الطبيعي لتعيين المنطقة الحرجة لدرجة Z، علي سبيل المثال، فإن القيمة الحرجة لِـ Z بمستوى دلالة 05. تكون 1.96 ±.

مثال تطبيقـي:

لتوضيح التقدير التقريبي الطبيعي، فإنه بإمكاننا التعامل مع نفس البيانات السابقة، حيث تهدف البيانات السابقة إلى مقارنة عينتين 1 و 2 مستخدمين عينتين مستقلتين يصل عدد كل منها إلي 6. وأن هذه البيانات قد ولدت قيمة تساوي ستة U=6 (لاحظ أنه عندما تكون العينات صغيرة، فإننا في العادة نستخدم جدول القيم الحرجة لاختبار مان- وتني بدلاً من جدول التقدير التقريبي الطبيعي، وعلى أية حال، فإننا نستخدم التقدير الطبيعي حتى نتمكن من مقارنة النتيجة المتحصل عليها بالنتيجة التي تحصلنا عليها من خلال اختبار مان- وتني الاعتيادي). وباستخدام المعادلة أعلاه، فإن درجة Z المناظرة لـِ: U=6 تكون:

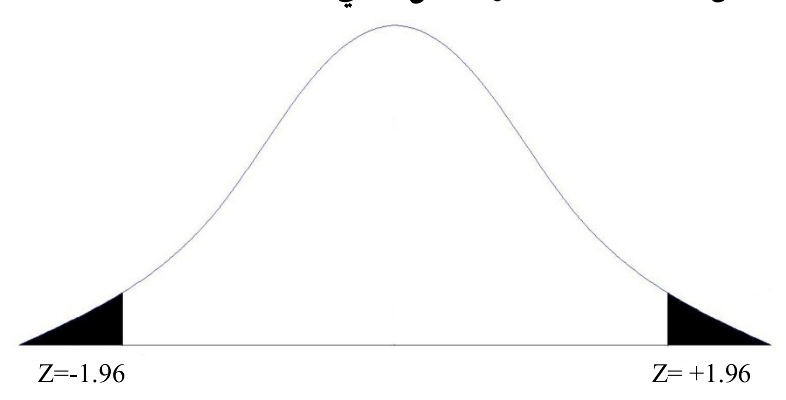
$$Z = \frac{u - n_1 n_2}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

$$= \frac{6 - (6)(6)}{\sqrt{\frac{(6((6)(6+6+1)}{12}}}$$

$$= \frac{-12}{\sqrt{\frac{468}{12}}}$$

$$= -1.92$$

وطبقاً لوحدة الجدول الطبيعي فإن 5 % المتطرفة للتوزيع الطبيعي تكون موجودة في الذيلين أبعد من Z=± 1.96 (انظر الشكل التالي:)



بدرجة دلالة 50. = ∝، و أن درجة Z أبعد من 1.96 ± فهي بالتالي تقود إلى رفض الفرق الصفري. إن القيم المحسوبة لـِ Z=-1.92 لا تقع في المنطقة الحرجة، وعليه فإن القرار الذي يمكن أن نصل إليه مفاده أنه ليس بإمكاننا رفض الفرض الصفري. (لاحظ أننا قد توصلنا إلى نفس النتيجة للاختبار الأصلي باستخدام القيمة الحرجة لجدول اختبار مان – وتني U).

إطار:

إعداد تقرير حول نتائج اختبار U :

بخلاف كثير من النتائج الإحصائية الأخرى، لا توجد قواعد صارمة لإعداد تقرير حول نتائج اختبار مان – وتني U. وبناءً على دليل الجمعية الأمريكية لعلم النفس APA فإنه يقترح أنه ينبغي في إعداد التقرير أن يشمل ملخصاً للبيانات (مثل: معلومات حول حجم العينة، ومجموع الرتب)، والإحصاء المتحصل عليه، وقيمة P. ففي المثال الذي بين أيدينا، فإن نتائج الدراسة يمكن أن تعد في شكل التقرير التالي:

الدرجات الأصلية قد تم ترتيبها، وتم حساب اختبار U لمقارنة الرتب لعدد 6 مشاركين في المعالجة الأولى (أ)، و 6 مشاركين في المعالجة الثانية (ب). وقد أشارت النتيجة إلى أنه يوجد فرق دال بين المعالجتين $U=6, P > 0.05$ لمجموع رتب مساوية لـِ 27 للمعالجة (أ)، و 51 للمعالجة (ب) (2).

الافتراضات والمحاذير لاستخدام اختبار مان ـ وتني U:

يعتبر اختبار مان – وتني U اختباراً مفيداً جداً كبديل لقياس اختبار T المستقل، باعتبار أن اختبار U لا يتطلب تجانسية التباين Homogeneity of Variance أو التوزيعات الطبيعية، ومن هنا يمكن استخدامه في المواقف التي لا يكون فيها استخدام الاختبار ممكناً، وعلى أية حال، فإن اختبار U لا يتطلب مشاهدات مستقلة Independent observations، وإنما يفترض أن يكون المتغير التابع متغيراً متصلاً (المتغير

المتصل هو ذلك المقياس الذي يمتلك عدداً لا متناهياً من النقاط الواضحة المعالم). إن أحد نتائج هذه الحقيقة أنه من غير المرجح لشخصين أن يمتلكا بالضبط نفس الدرجة. وهذا يعني أنه يجب أن يوجد قليل من الدرجات المتعادلة في البيانات. وعندما تكون لدينا بيانات عينة لديها عدة درجات متعادلة، فهذا يعني أن الافتراض الأساسي لاختبار مان- وتني U قد انتهك. وفي هذه الحالة ينبغي على الباحث أن يكون حذراً في استخدامه لاختبار مان - وتني U. وتجدر الإشارة هنا إلى أنه إذا كان هناك قليل من الدرجات المتعادلة نسبياً يمكن للباحث استخدام اختبار U، لكن ينبغي عليه اتباع الإجراء المعياري لترتيب الدرجة المتعادلة [4].

ثانياً: اختبار الإشارة والرتب ولكوكسن (T) Wilcoxon Test:

لقد صُمِّمَ هذا الاختبار لتقييم الفرق بين معالجتين باستخدام بيانات جاءت من تجربة متكررة القياس a repeated - measures. لاحظ أن دراسة القياسات المتكررة يتعلق فقط بعينة واحدة حيث إن كل فرد في العينة يتم قياسه مرتين: مرة في المعالجة الأولى، والمرة الأخرى في المعالجة الثانية. والفرق بين القياسين يتم تسجيله كدرجة تتعلق بذلك الفرد.

ويتطلب اختبار ولكوكسن ترتيب الفرق بشكل منظم من الأصغر إلى الأكبر From Smalest to Largest بغض النظر عن الإشارة أو الاتجاه. علي سبيل المثال، في الجدول الموالي (أ) يبين لنا الفروق غير العددية Nonnumerical differences والرتب Ranks لعينة تحتوي على أربعة مشاركين N=4. فإذا كانت الفروق تمثل قيماً عددية Numerical Values، فيمكن أيضاً ترتيبها في إطار حجمها المطلق absolute magnitude. أما الجدول (ب) فيوضح لنا أن N=5 مع فرق الدرجات العددية وترتيبها.

جدول (أ) فرق الدرجات في الجزء (أ) الفروق يتم قياسها كزيادة نسبية أو نقصان، والقيم غير العددية تم ترتيبها طبقاً للحجم، مستقلة في الاتجاه.

أما الجزء (ب) فإن فرق الدرجات يمثل قيماً عددية، والفرق تم ترتيبه مستقلاً في الاتجاه.

(أ)

المشاركون	الفرق من المعالجة الأولى للمعالجة الثانية	الرتبة
أ	زيادة صغيرة	2
ب	نقص كبير جداً	4
ج	زيادة متوسطة	3
د	زيادة صغيرة جداً	1

(ب)

المشاركون	فرق الدرجة	الرتبة	
أ	+ 4	2	$+\sum R=6$
ب	- 14	5	
ج	+ 9	4	$-\sum R=9$
د	- 1	1	
هـ	- 6	3	T = 6

المصدر: frederick J.Gravetter and Larry B.wallnau statistics for the Behavioral sciences , 8th ed ,wadsworth cengage learning,USA ,2010,p 675.

الفروض المتعلقة باختبار ولكوكسن :T

يصاغ الفرض الصفري لاختبار ولكوكسن ببساطة، أنه لا يوجد فرق ثابت ونظامي بين المعالجتين.

H_O: لا يوجد فرق بين المعالجتين. وعليه، فإنه في المجتمع العام لا يوجد اتجاه في الفروق في الدرجات إما أن تكون بشكل نظامي موجبة أو بشكل نظامي سالبة.

H_i: يوجد فرق بين المعالجتين. وعليه، فإنه في المجتمع العام يكون اتجاه الفروق في الدرجات إما بشكل نظامي موجب أو بشكل نظامي سالب.

فإذا كان الفرض الصفري (H_O) فرضاً صحيحاً، فإن أيّاً من الفروق التي توجد في بيانات العينة يجب أن يكون مردها لعامل الصدفة، وعليه، فإننا بذلك سنتوقع فروقاً موجبةً وفروقاً سالبةً تكون متمازجة بالتساوي. intermixed evenly. وعلى الجانب الآخر، إذا كان الفرق في الدرجات فرقاً ثابتاً ونظامياً بين المعالجتين، فإنه بالتالي يجب أن يتسبب في وجود درجات كبيرة في إحدى هذه المعالجات مقارنة بالدرجات في المعالجة الأخرى.

وهذا يجب أن يقود إلى توليد فروق في الدرجات التي تتجه بشكل ثابت نحو الموجب أو السالب. ويستخدم اختبار ولكوكسن T الإشارات (S) والرتب RANKS في الفرق في الدرجات لتقييم ما إذا كان هناك فرقُ ذو دلالة إحصائية بين المعالجتين[6].

حساب وتفسير اختبار ولكوكسن (T):

كما هو الحال في معظم الاختبارات غير البارامترية، فإن حساب اختبار ولكوكسن يكون إلى حد بعيد بسيطاً في عملية حسابه حيث:

1- ترتيب القيم المطلقة لفرق الدرجات.

2- تقسيم الترتيب إلى مجموعتين: مجموعة مرتبطة بالفروق الموجبة (الزيادة) والأخرى مرتبطة بالفروق السالبة (النقصان).

3- إيجاد مجموعة الرتب لكل مجموعة، واعتماد المجموع الأصغر لهاتين المجموعتين كإحصائي الاختبار لاختبار ولكوكسن. ويشار إلى ذلك بالرمز T.

ففي البيانات في الجدول السابق (أ)، على سبيل المثال، فإن الزيادة في الترتيب كانت متعلقة بـِ: 1، 2، و 3 التي يصل مجموعتها إلى $\sum R=6$. في حين يوجد نقص واحد فقط المرتبط بالترتيب 4. ومن هنا فإن قيمة T لمجموعة هذه البيانات: T=4. أما في الجدول (ب)، فإن الفرق في الدرجات (الدرجات الموجبة) فهي مرتبطة بالترتيب 2، و4 التي يصل مجموعها إلى: $\sum R=6$. وأن فرق الدرجات التي تحمل إشارة سالبة فهي مرتبطة بـِ: 1،3، و5 التي يصل مجموعها إلى $\sum R=9$، وأن قيمة T لهذه الدرجات وصلت إلى: T= 6

وكما بينا سابقاً، فإن أقوى تأثير معالجة يجب أن يسبب الفرق في الدرجات بحيث يكون فرقاً موجباً أو فرقاً سالباً. وأنه في الحالة المتطرفة فإن كل الفروق تكون في نفس الاتجاه مولدة T=0. فعلى سبيل المثال، عندما تكون كل الفروق فروقاً موجبة، فإن مجموع الرتب السالبة تكون صفراً. وعلى الجانب الآخر، إذا لم يكن هناك أي تأثير للمعالجة N0 treatment effect. فإن الإشارات المتعلقة بالفرق في الدرجات يجب أن تتمازج بالتساوي. وفي هذه الحالة يكون اختبار ولكوكسن T كبيراً جداً نسبياً. وبشكل عام، فإن قيمة T الصغيرة (قريبة من الصفر) تقدم دليلاً للفرق الحقيقي بين حالتي المعالجة. إن توزيع كل القيم المحتملة لِـ: T قد تم بناؤه. وأن القيم الحرجة لمستوي دلالة (α) 0.05 ومستوى دلالة 0.01 يمكن الحصول عليها من الجدول المتعلق بقيم T الحرجة (انظر الملحق رقم 1) وكلما ولدت بيانات عينة أقل أو مساوية للقيمة الحرجة، فإن ذلك يقود إلى رفض الفرض الصفري Ho . ولمقاييس الدراسة المتكررة لعينة تصل إلى N=15.15، علي سبيل المثال، فإن الجدول المتعلق باختبار ولكوكسن للقيم الحرجة T=25 اختبار ثنائي الجانب على مستوى دلالة 05. = α. وهذا يعني أن قيمة T الحرجة التي تساوي 25 أو أصغر تكون أكثر ترجيحاً أن تحدث (احتمالية 05. أو أقل) إذا كان الفرض الصفري صحيحاً[7].

الدرجات المتعادلة ودرجات صفر:

بالرغم من أن اختبار ولكوكسن T لا يتطلب حسابه التوزيعات الطبيعية Normal distributions، إلا أنه يفترض أن يكون المتغير التابع متغيراً متصلاً. وكما نوهنا سلفاً عند الحديث عن اختبار مان- وتنى، فإن هذا الافتراض ينطوي علي أن تلك الدرجات المتعادلة ينبغي أن تكون إلى حد كبير بعيدة الاحتمال. فعندما يظهر التعادل في بيانات العينة، ينبغي على الباحث أن يكون قلقاً بأن فرضية المتصلية Continuity قد تم انتهاكها. ومن هنا قد لا يكون اختبار ولكوكسن T اختباراً مناسباً.

أما إذا كان هناك تعادل قليلُ نسبياً في البيانات، فإن معظم الباحثين يفترضون أن البيانات هي حقاً بيانات متصلة ولكن قد تم قياسها بشكل فج، في هذه الحالة، يمكن

للباحث استخدام اختبار ولكوكسن T و إبلاء القيم المتعادلة انتباهاً خاصاً عند العمليات الحسابية[8].

تجدر الإشارة إلى أن هناك نمطين مختلفين للدرجات المتعادلة:

1- إن الموضوع الخاضع للدراسة قد تكون درجته متساوية في المعالجة الأولى وفي المعالجة الثانية، وينتج عن ذلك فرق يساوي صفراً (0).

2- قد يكون لدينا موضعان أو أكثر متماثلان في فرق الدرجات (متجاهلين علامة هذا الفرق)[9].

وعندما تحتوي البيانات أفراداً بفروق درجات تساوي صفراً (0)، حينئذٍ ينبغي طرح أولئك الأفراد جانباً من التحليل، وخفض حجم العينة المدروسة (N) ومع ذلك، فإن هذا الإجراء يتجاهل حقيقة أن فرق الدرجة التي تساوي صفراً (0) يكون دليلاً يبقى على الفرضية الصفرية (Ho). إن أفضل إجراء يمكن اتباعه هو تقسيم فروق الدرجات التي تساوي صفراً بشكل متساوٍ بين الدرجات الموجبة والدرجات السالبة (إذا كان لدينا رقم مفرد لفروق الدرجات التي تساوي صفراً 0). فإنه ينبغي علينا أن نتجاهل أحدهما ويقسم الباقي بشكل متساوٍ) إن هذا الإجراء الثاني يهدف إلى زيادة $\sum R$ لكل من الرتب الموجبة والرتب السالبة التي تزيد من القيمة النهائية (T)، وتجعل هذه القيمة أكثر ترجيحاً على أنه سوف نبقي على الفرض الصفري.

وعندما تكون لدينا درجات متعادلة بين فروق الدرجات، فإن كل واحدة من الدرجات المتعادلة ينبغي أن يخصص لها الوسط الحسابي للرتب المتعادلة.

اختبار الفرض باستخدام ولكوكسن T:

المثال الموالي يعطينا مثالاً بيّناً لاختبار ولكوسن باستخدام بيانات لنمطين من درجات متعادلة.

مثال: البيانات التالية تمثل عشر هيئات. وكل هيئة من هذه الهيئات كانت تهدف إلي زيادة نسبة العاملين لديها الذين يشاركون في برنامج للتبرع بالدم. البيانات التالية توضح

نسبة المشاركين في البرنامج في السنة الماضية (قبل الشروع في الدعاية لهذا البرنامج). وخلال هذا العام وباستخدام اختبار ولكوكسن T لتقييم ما إذا كانت البيانات تقدم دليلاً كافياً فإن الدعاية لهذا البرنامج قد أحدثت تأثيراً دالاً على عملية التبرع بالدم. لقد رُتِّبَتْ بيانات الجدول بشكل منظم وذلك طبقاً للقيمة المطلقة لفرق الدرجات.

نسبة المشاركين					
الهيئة	قبل الدعاية	بعد الدعاية	الفرق	التركيب تجاهل الأصفار	التركيب تضمين الأصفار
1	18	18	0	–	1.5
2	24	24	0	–	1.5
3	31	30	−1	1	3
4	28	24	−4	2	4
5	17	24	+7	3	5
6	16	24	+8	4	6
7	15	26	+11	5.5	7.5
8	18	29	+11	5.5	7.5
9	20	36	+16	7	9
10	9	28	+19	8	10

المصدر: Ibid , P. 677.

أ- تجاهل الفروق التي تساوي صفرا (0):

إن أولى الخطوات لإجراء اختبار ولكوكسن T هي الأخذ بالتوصية المتعلقة بتجاهل الفروق التي تساوي صفراً.

الخطوة الأولى: صياغة الفرض الصفري:

H_O: لا يوجد تأثير للدعاية لهذا البرنامج، عليه، فإن أية فروق في الدرجات يكون مرده لعامل الصدفة. وهنا ينبغي ألا يكون هناك نمطُ ثابت لهذه الفروق.

الخطوة الثانية: إن الهيئتين اللَّتَيْن لديهما فرق في الدرجات مساوياً لصفر ينبغي تجاهلها. وأن حجم العينة بالتالي ينخفض إلى: N=8 ودرجة الدلالة .05 = α. والقيمة الحرجة لاختبار ولكوكسن T يساوي 3.

إن أي قيمة لإحصائي الاختبار أقل أو مساوية لـ: 3 تقود إلى رفض الفرض الصفري.

الخطوة الثالثة: **تحدد مجموع الفروق ذات الرتب الموجبة** (+):

3 ، 4 ، 5 ، 5 ، 5 ، 5 ، 7 ، 8 . $\sum R+ = 33$

ومجموع الفروق ذات الرتب السالبة (–):

1 ، 2 ، . $\sum R - = 3$

بعد ذلك يتم اعتماد أصغرهما. عليه: T=3

الخطوة الرابعة: لما كانت قيمة T=3 أصغر من القيمة الحرجة لـ:T؛ عليه نرفض الفرق الصفري ونخلص إلى القول بأنه يوجد تغيير دال في المشاركة بالتبرع بالدم في حملة الدعاية التي قام بها الهلال الأحمر [10].

بـ تضمين الفروق التي تساوي صفرا (0):

إذا أضفنا الشخصين اللَّذَيْن لديهما فرق الدرجات المساوية لصفر، حينئذٍ يصبح حجم العينة N=10 بدرجة دلالة 0.55 = α والقيمة الحرجة لاختبار ولكوكسن T هي 8. ومهما كانت فروق الدرجات التي تساوي صفراً متعادلةً من حيث الترتيب الأول والثاني، فقط أَعْطِ لكل منهما 1.5 وأن أحد هذه الرتب تم تخصيصها للمجموعة الموجبة، و الأخرى للمجموعة السالبة كما هو مبين في النتيجة التالية:

$$\sum R+ = 1.5 + 5 + 6 + 7.5 + 7.5 + 9 + 10$$

$$= 46.5$$

و

$$\sum R- = 1.5 + 3 + 4$$

$$= 8.5$$

ولما كانت قيمة ولكوكسن T أصغر مجموع لهاتين القيمتين T=8.5 ولما كانت هذه القيمة أكبر من القيمة الحرجة، عليه نقبل الفرض الصفري.

ونَخْلُصُ إلي القول بأن هذه البيانات لم تقدم لنا دليلا كافياً على أن هناك تغيراً دالاً لدى المشاركين في برنامج التبرع بالدم.

تجدر الإشارات هنا إلي أنه بإضافة فروق الدرجات التي تساوي صفراً (0) للاختبار قد غيرت من النتيجة الإحصائية. فالفروق في الدرجات التي تساوي صفراً (0) هي في الواقع تؤشر بأن الفرض الصفري فرضُ صحيحُ.

التقدير التقريبي الطبيعي لاختبار ولكوكسن (T): [11]

عندما تكون العينة كبيرة نسبياً، فإن قيم إحصاء ولكوكسن T تميل إلي تشكيل توزيع طبيعي.في هذه الحالة، فإن من الممكن إجراء الاختبار وذلك باستخدام إحصاء درجة Z، والتوزيع الطبيعي،بدلاً من النظر إلي قيمة T في جدول ولكوكسن. وعندما يكون حجم العينة أكبر من 20، فإن التقدير التقريبي الطبيعي يكون دقيقاً وأنه بالإمكان استخدامه.ففي العينات الكبيرة التي تتعدي N=>50، فإن الجدول المتعلق بدرجات ولكوكسن لا يمدنا علي نحو نموذجي بأية قيم حرجة.عليه، يصبح من الضرورة بمكان استخدام التقدير التقريبي الطبيعي.

إن الإجراء المتبع للتقدير التقريبي لاختبار ولكوكسن (T) هو كالتالي:

1- إيجاد مجموع الرتب ذات الإشارات الموجبة (+) وإجمالي الرتب التي تحمل الإشارات السالبة (-) كما بينا سابقاً. وأن قيمة ولكوكسن (T) تمثل أصغر القيمتين.

2- عندما تكون (N) أكبر من 20،فإن قيم ولكوكسن (T) تشكل توزيعاً طبيعياً بمتوسط:

$$U = \frac{n(n+1)}{4}$$

وانحراف معياري:

$$\sigma = \sqrt{\frac{n(n+1)(2N+1)}{24}}$$

في هذه الحالة يتم تحويل ولكوكسن (T) إلي درجة Z :

$$Z=\frac{x-u}{\sigma}=\frac{\frac{T-N(n+1)}{4}}{\sqrt{\frac{n(n+1)(2N+1)}{24}}}$$

3- يستخدم جدول المنحي الطبيعي المعياري لتحديد المنطقة الحرجة لدرجة Z. علي سبيل المثال، القيم الحرجة تكون 1.96± بدرجة دلالة 05. = α.

مثال تطبيقي:

كما أشرنا في موضعه إلى أن التقدير التقريبي لاختبار (T) لعينة أكبر من 20 فرداً، أنه سوف تبرهن لنا العمليات الحسابية لنفس بيانات المثال المنفرط الذي قدمناه كمثال لحساب (T).

أ- تجاهل الفروق التي تساوي صفرا (0):

عندما تم تجاهل الفروق التي تساوي صفراً (0) لتلك الهيئتين، قد تحصلنا علي N=8 و T=3. وباستخدام التقدير التقريبي الطبيعي فإن تلك القيم تولد:

$$u=\frac{(n+1)}{4}=\frac{8(9)}{4}=18$$

$$\sigma=\sqrt{\frac{n(n+1)(2n+1)}{24}}$$

$$=\sqrt{\frac{8(9)(17)}{24}}=\sqrt{51}=7.14$$

وبهذه القيم فإن T التي تحصلنا عليها تقابل درجة Z التالية:

$$Z=\frac{T-u}{\sigma}=\frac{3-18}{7.4}=\frac{-15}{7.14}=-2.10$$

ومع الحدود الحرجة 1.96 ±، فإن درجة Z المتحصل عليها تكون قريبة من الحد، ولكنها كافية لأن تكون دالة على مستوى 05. = α. لاحظ أن هذه النتيجة مساوية

بالضبط لتلك النتيجة التي توصلنا إليها باستخدام جدول توزيع ولكوكسن T عندما تجاهلنا الفروق التي تساوي صفراً (0) في المثال المنفرط.

تضمين الفروق التي تساوي صفرا (0):

إذا أضفنا الشخصين اللّذَيْن لديهما فرق الدرجات المساوية لصفر (0)؛ حينئذٍ يصبح حجم العينة N = 10 و T = 8.5. وباستخدام التقدير التقريبي الطبيعي، فإن هذه القيم تولد:

$$u = \frac{(n+1)}{4} = \frac{10(11)}{4} = 27.5$$

$$\sigma = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

$$= \sqrt{\frac{10(11)(21)}{24}} = \sqrt{96.25} = 9.81$$

وبهذه القيم فإن قيمة T المتحصلين عليها تماثل درجة Z التالية:

$$Z = \frac{x-u}{\sigma} = \frac{8.5-27.5}{8.81} = \frac{-19}{9.81} = -1.94$$

ومع الحدود الحرجة 1.96 ±، فإن درجة Z التي تحصلنا عليها تكون قريبة من الحد، ولكنها ليست كافية لأن تكون دالة على مستوى 05. = α. مرة ثانية، إن هذه النتيجة هي تلك النتيجة نفسها التي توصلنا إليها مستخدمين جدول T، عند تضميننا للفروق التي تساوي صفراً (0) في المثال السابق.

ولإعداد تقرير حول النتائج المتوصل إليها من خلال استخدام هذا الاختبار (T)، يمكننا القول بإنه لا يوجد شكل محدد لإعداد تقرير حول نتائج اختبار ولكوكسن (T)، فإنه يقترح أن يشتمل هذا التقرير على ملخص للبيانات والقيمة المتحصل عليها من إحصاء هذا الاختبار. إضافة إلى قيمة P. كذلك يقترح أن يصف التقرير المعالجة التي تمت بخصوص فروق الدرجات التي تساوي صفراً (0) في التحليل. ففي الدراسة التي أوردناها في هذا الجزء من هذا الفصل، بينت أن هناك هيئتين لا تَغَيُّرَ في مشاركتهما في حملة التبرع

بالدم؛ وبالتالي قد تم تجاهلهما قبل عملية التحليل. وبالتالي تم ترتيب ثمان مجموعات فقط (N=8) وفقاً للمتغير في مستوى المشاركة. وقد تم استخدام اختبار T لتقييم هذه البيانات. وقد بَيَّنَتِ النتائج أن هناك زيادةً دالةً في المشاركة بعد الحملة الدعائية: $T = 3, p < .05$ للرتب بزيادة تصل إلى 33، وللرتب بالنقصان تصل إلى 3.

ثالثاً: اختبار كروسكال وليز Kruskal - Wallis Test:

يستخدم اختبار كروسكال - وليز لتقييم الفروق بين ثلاث عينات أو أكثر مستخدماً بيانات جاءت من خلال تصميم المقاييس المستقلة. ويعتبر هذا المقياس مقياساً بديلاً لتحليل التباين أحادي الجانب ANOVA.

وإذا كان مقياس أنوفا يتطلب إجراؤه درجات عددية يمكن استخدامها لحساب المتوسطات والتباين؛ فإن اختبار كروسكال - وليز على الجانب الآخر، يتطلب، ببساطة، ترتيب الأفراد بشكل منظم ترتيباً تصاعدياً من أصغر رتبة حتى أكبرها على المتغير المطلوب قياسه. وتجدر الإشارة هنا إلى أن اختبار كروسكال - وليز شبيه لاختبار مان - وتني الذي تناولناه في بداية هذا الفصل. إلا أن الأخير يستخدم للمقارنة بين مجموعتين، في حين أن اختبار كروسكال - وليز يستخدم في تقييم الفروق بين ثلاث مجموعات أو أكثر. إضافة إلى أن مقياس كروسكال - وليز يمكن استخدامه إذا كانت البيانات الأصلية العددية قد تحولت إلى قيم ترتيبية. المثال التالي يوضح العملية التي من خلالها يتم تحويل الدرجات العددية إلى رتب يمكن استخدامها في تحليل كروسكال - وليز.

مثال: يوضح المثال التالي البيانات الأصلية لدراسة تتعلق بثلاث مجموعات. ولأجل إعداد البيانات بشكل يمكن استخدامه في اختبار كروسكال - وليز، يتطلب بادئ ذي بدء ترتيب كل الدرجات الأصلية وفقاً للطريقة المعتادة لترتيب الدرجات المتعادلة؛ أي أن كل الدرجات الأصلية بعد ذلك تستبدل برتبها.

جدول رقم (20ـ 2)

البيانات المعدة للتحليل باستخدام اختبار كروسكال ـ وليز. البيانات الأصلية تحتوي على درجات عددية كما هو مبين في الجدول (أ)، والبيانات الأصلية سجلت بشكل منظم وخصصت لها رتب، والرتب أصبحت بديلاً للدرجات الأصلية كما هو مبين في الجدول (ب)

البيانات الأصلية للدرجات العددية (أ)

1	3	3
14	2	26
3	14	8
21	9	14
5	12	19
16	5	20

إيجاد ترتيب كل درجة

الدرجات الأصلية العددية	الترتيب الأصلي
2	1
3	2
5	3.5
5	3.5
8	5
9	6
12	7
14	9
14	9
14	9
16	11
19	12
20	13
21	14
26	15

رتب البيانات الترتيبية (ب)

1	2	3
9	1	15
2	9	5
14	6	9
3.5	7	12
11	3.5	13

المصدر: Frederick J. Gravetter and Larry B. Wallnau , op.cit , P. 682.

الفرضية الصفرية لاختبار كروسكال- وليز:

كما هو الحال في كل الاختبارات المتعلقة بالبيانات الترتيبية، فإن الفرضية لاختبار كروسكال - وليز تميل إلى شيء من الغموض. وبشكل عام، يصاغ الفرض الصفري لاختبار كروسكال - وليز بأنه لا توجد فروق ذات دلالة إحصائية بين المجموعات الخاضعة للمقارنة. وبشكل أكثر تحديداً، يصاغ الفرض الصفري (Ho) أنه لا يوجد ميل بأن رتب أي مجموعة ستكون بشكل منظم أعلى (أو أقل) من رتب أي مجموعة أخرى. عليه، يمكن صياغة الفرض الصفري والفرض البديل بالشكل التالي:

Ho: لا يوجد ميل للرتب في أي مجموعة أن تكون بشكل منظم أعلى أو أقل من رتب أية مجموعة أخرى.

Hi: أن الرتب على الأقل في واحدة من المجموعات تكون بشكل منظم أعلى أو أقل من رتب في مجموعة أخرى. وبالتالي توجد فروق بين المجموعات.

المعادلات المرتبطة باختبار كروسكال- وليز:

الجدول التالي يبين لنا تحويل البيانات من بيانات عددية إلى رتب. ويتطلب الأمر عند تطبيق المعادلات المتعلقة بهذا الاختبار اتباع الإجراءات التالية:

1- تضاف الرتب إلى بعضها البعض لكل مجموعة للحصول على العدد الإجمالي أو قيمة T لحالة المعالجة. وأن هذه القيمة المتحصل عليها (T) يتم استخدامها في معادلة كروسكال - وليز.

2- عدد الحالات لكل مجموعة يشار إليها بـِ (n) الصغيرة.

3- العدد الإجمالي للمجموعات بكاملها يشار إليه بـِ (N) الكبيرة.

وتولد لنا معادلة كروسكال - وليز إحصاءً يشار إليه بـِ (H) ولديه تقريباً نفس توزيع مربع كاي (X^2) بدرجة حرية تعرف بـِ: عدد المجموعات ناقص واحد.

الجدول (20-3) التالي:

1	2	3
9	1	15
2	9	5
14	6	9
3.5	7	12
11	3.5	13
$T_1 = 39.5$	$T_2 = 26.5$	$T_3 = 54$
$n_1 = 5$	$n_2 = 5$	$n_3 = 5$

وللبيانات الواردة في الجدول أعلاه، فإن عدد الحالات 3، وبالتالي فإن معادلة مربع كاي تولد لنا قيمة متعلقة بمربع كاي بدرجة حرية 2.

والمعادلة المتعلقة بإحصاء كروسكال - وليز هي كالتالي:

$$H = \frac{12}{N(N+1)}\left(\sum \frac{T^2}{n}\right) - 3(N+1)$$

وباستخدام البيانات السابقة، فإن معادلة كروسكال - وليز تولد لنا قيمة مربع كاي:

$$H = \frac{12}{15(16)}\left(\frac{(39.5)^2}{5} + \frac{(26.5)^2}{5} + \frac{(54)^2}{5}\right) - 3(16)$$

$$= 0.05\ (312.05 + 140.45 + 583.2) - 48$$

$$= 51.785 - 48$$

$$= 3.785$$

وبدرجة حرية (2) df=2، فإن القيمة الحرجة لمربع كاي تساوي 5.99 بدرجة دلالة $\alpha = .05$. ولما كانت القيمة المحسوبة لمربع كاي (3.785) أصغر من القيمة الجدولية؛ فإننا بالتالي نكون غير قادرين على رفض الفرض الصفري (Ho) حيث إن البيانات لم تقدم لنا دليلاً كافياً يمكننا من القول بأنه توجد فروق دالة بين هذه المجموعات الثلاثة.

ولإعداد تقرير حول نتائج اختبار (H)، يمكننا القول بأنه بعد ترتيب درجات الأفراد قد تم استخدام اختبار كروسكال - وليز لتقييم الفروق بين المجموعات الثلاثة. وقد جاءت النتيجة لتشير إلى أنه لا توجد فروق دالة بين هذه المجموعات.

(12) $H: 3.785(df2, N=15) P > .05$

رابعاً: اختبار فريدمان: كبديل لمقاييس أنوفا المتكررة The Friedman Test:

يستخدم اختبار فريدمان لتقييم الفروق بين ثلاث أو أربع مجموعات مستخدماً بيانات لتصميم مقاييس متكررة.

ويعتبر اختبار فريدمان بديلاً لمقاييس انوفا المتكررة التي تحدثنا عنها في الفصول السابقة. ومع أن تحليل أنوفا يتطلب درجات عددية Numerical Scores التي يمكن استخدامها في حساب المتوسطات والتباين، إلا أن اختبار فريدمان يتطلب ببساطة أن يكون الباحث قادراً على ترتيب المشاركين على المتغير الذي يتم قياسه. كذلك نجد أن مقياس فريدمان مشابه لاختبار ولكوكسن. حيث إن الأخير - كما أشرنا سابقاً - يتعلق بمقارنة حالتين فقط. في حين أن اختبار فريدمان يستخدم للمقارنة بين ثلاث أو أكثر من الحالات.

مستوى البيانات المطلوبة لاختبار فريدمان:

لإجراء اختبار فريدمان، يتطلب الأمر وجود عينة واحدة One Sample لكل فرد مشارك في كل حالات المعالجة المختلفة. ويتطلب هذا الاختبار ترتيب كل حالات المعالجة لكل فرد مشارك في الدراسة. كما يمكن استخدام اختبار فريدمان إذا كانت البيانات الأصلية تحتوي على درجات عددية، ومع ذلك ينبغي تحويل هذه الدرجات العددية إلى رتب قبل إجراء هذا الاختبار.

والمثال التالي يبين طريقة استخدام هذا الاختبار.

مثال: عينة تتكون من خمسة مشاركين N = 5، تم اختبارهم تحت أربعة أنواع من

الدواء. وقد كانت الدرجات الأصلية تحتوي على درجات عددية (انظر الجدول "أ"). ولكي يطبق اختبار فريدمان فقد حُولت هذه الدرجات العددية إلى رتب 1، 2، 3 و 4 لتطابق حجم الدرجات الأصلية. فعلى سبيل المثال، أن المشارك رقم (1) لديه درجات 3، 4، 6، 7. ولهذا المشارك $X = 3$ وهي أصغر درجة. وبالتالي كان ترتيبه (1)؛ و $X = 4$ هي الدرجة الثانية من حيث الصغر؛ وبالتالي كان ترتيبه (2)، أما الدرجتان الأخريان فقط تحصلتا على ترتيب (3) و (4). في حين أن المشارك رقم (2) لديه درجات متعادلة لِـ: $X = 3$ للدواء (أ) والدواء (ب)؛ وبالتالي قد تحصلا على ترتيب (2.5). (2+3 = 5/ 2 تساوي 2.5). إن مجموعة الرتب كاملة يبينها الجزء من الجدول (20 – ب).

(أ): جدول (20 ـ أ) الدرجات الأصلية للدواء

المشارك	دواء بدون مفعول (مهدئ للمريض)	عقار (أ)	عقار (ب)	عقار (ج)
1	3	4	6	7
2	0	3	3	6
3	2	1	4	5
4	0	1	3	4
5	0	1	4	3

(ب): جدول (20 ـ ب) ترتيب حالات المعالجة بشكل منظم لكل مشارك

المشارك	دواء بدون مفعول (مهدئ للمريض)	عقار له مفعول (أ)	عقار له مفعول (ب)	عقار له مفعول (ج)
1	1	2	3	4
2	1	2.5	2.5	4
3	2	1	3	4
4	1	2	3	4
5	1	2	4	3
	$R_1 = 6$	$R_2 = 9.5$	$R_3 = 15.5$	$R_4 = 19$

المصدر: Frederick J. GRavetter and Larry B. wallnau , op.cit , p.685.

الفرض الصفري لاختبار فريدمان: [13]

بشكل عام، فإن الفرض الصفري لاختبار فريدمان يصاغ بالشكل التالي:

لا توجد فروق بين حالات المعالجة الخاضعة للمقارنة. فإذا كان الفرض الصفري فرضاً صحيحاً، فإن الرتب يجب أن تتوزع بشكل عشوائي؛ وبالتالي لا توجد أية نزعة بأن هذه الرتب لإحدى المعالجات تكون بشكل نظامي أعلى (أو أقل) من الرتب في المعالجات الأخرى. وعليه، فإن الفروض المتعلقة باختبار فريدمان يمكن أن تصاغ بالشكل التالي:

Ho: لا يوجد فرق بين المعالجات. وعليه فإن الرتب في إحدى حالات المعالجة لا يجب أن تكون أعلى أو أقل من الرتب في أية معالجة أخرى.

Hi: يوجد فروق بين المعالجات، وعليه، فإن الرتب على الأقل في إحدى حالات المعالجة يجب أن تكون بشكل منظم أعلى أو أقل من الرتب في معالجة أخرى.

معادلة وحساب اختبار فريدمان:

لحساب اختبار فريدمان ينبغي للباحث اتباع الخطوات التالية:

1- ترتيب درجة كل مشارك من المشاركين الخمسة.

2- تجمع الرتب لكل فرد مشارك كما هو مبين في السطر الأخير من الجدول. فإذا كان الفرض الصفري فرضاً صحيحاً، فإن هذه المجاميع يجب أن تكون تقريباً بنفس الحجم. لأنه لا أحد من هؤلاء المشاركين لديه ترتيب أعلى أو أدنى من المشارك الآخر. وعلى الجانب الآخر، إذا ما وجد فرق ثابت بين هؤلاء المشاركين، إذاً على الأقل واحد من هذه المجاميع يجب أن تكون لافتة للنظر بأنها أكبر (أو أصغر) من مجموع مشارك أخرى.

ويشار إلى مجموع الرتب لكل مشارك بالرمز R_1، R_2 وهكذا... ففي الجدول (ب) نجد أن المعالجة الأولى (الدواء المهدئ Placebo) يصل مجموعها إلى $R_1 = 6$ في حين يصل مجموع المعالجة الثانية $R_2 = 9.5$، والمعالجة الثالثة $R_3 = 15.5$، في حين يصل مجموع المعالجة الأخيرة إلى: $R_4 = 19$.

أما باقي الرموز المتعلقة باختبار فريدمان تحتوي على n و K والتي تقابل عدد الأفراد في العينة وعدد حالات المعالجة على التوالي. ففي البيانات الواردة في جدول (ب) فإن: N = 5 و K = 4.

ولما كان اختبار فريدمان اختباراً لتقييم الفروق بين المعالجات من خلال حساب إحصائي الاختبار التالي:

$$X_r^2 = \frac{12}{nk(k+1)}\sum R^2 - 3n(k+1)$$

لا حظ أن الإحصاء تم تعيّنه كمربع كاي (X^2) مقروناً برمز سفلي (r) ويقابل إحصاء مربع كاي (X^2) للرتب.

إن إحصاء مربع كاي لديه درجات حرية تحدد من خلال: df=k-1 ويقيِّم من خلال القيم الحرجة في توزيعات مربع كاي. وباستخدامنا للبيانات الواردة في الجدول رقم (ب)، فإننا سنتحصل على الإحصاء التالي:

$$X_r^2 = \frac{12}{nk(k+1)}\sum R^2 - 3n(k+1)$$

$$X_r^2 = \frac{12}{5(4)(5)}\left((6)^2 + (9.5)^2 + (15.5)^2 + (19)^2\right) - 3(5)(5)$$

$$= \frac{12}{100}(36 + 90.25 + 240.25 + 361) - 75$$

$$= 0.12\ (727.5) - 75$$

$$= 12.3$$

وبدرجة حرية df=k-1=3، وقيمة حرجة لمربع كاي تساوي 9.35. عليه فإن القرار الذي نصل إليه هو رفض الفرض الصفري، وبالتالي نصل إلى نتيجة مفادها أنه توجد فروق ذات دلالة إحصائية بين أحوال المعالجات الأربع.

لاحظ أنه في اختبار فريدمان ببساطة يمكننا القول بأنه توجد فروق دالة، في حين يمكننا القول عند التعامل مع اختبار أنوفا أنه توجد دلالة في فروق المتوسط. ولتفسير

نتائج اختبار فريدمان - مثله مثل كل الاختبارات ذات البيانات الترتيبية - لا يوجد شكل عام لإعداد تقرير حول النتائج المتحصل عليها من خلال هذا الاختبار؛ ومع ذلك، فإنه في إعداد التقرير المتعلق به ينبغي أن يحتوي على: القيمة المتحصل عليها بالنسبة لإحصاء مربع كاي، إضافة إلى القيم بالنسبة لدرجة الحرية df و N و P. بالنسبة للبيانات التي بين أيدينا التي تم البرهنة بها على هذا الاختبار، فالتقرير المعد ينبغي أن يدور حول النتائج التالية:

- بعد ترتيب البيانات الأصلية، فقد تم استخدام اختبار فريدمان لتقييم الفروق بين أربع حالات معالجة. وقد أشارت النتيجة أن هناك فروقاً دالة:

$$X_r^2 = (12.3(df = 3, n = 5), P < .05)$$

جدول (20ـ3) ملخص: مقارنة بين هذه المقاييس الأربعة

الاختبار	نمط العلاقات المختبرة	مستوى البيانات المطلوبة	تصميم الدراسة
مان-وتني U	تقييم الفرق بين حالتين	بيانات ترتيبية	بين الموضوعات: • عينات مستقلة • موضوع فردي
ولكوكسن (T)	تقييم الفرق بين حالتين	بيانات ترتيبية	داخل الموضوعات: • المقاييس المتكررة • الأزواج المتماثلة
كروسكال-وليز (H)	تقييم الفروق بين ثلاث حالات أو أكثر	بيانات ترتيبية	بين الموضوعات: • تصميمات غير مترابطة • (موضوعات مستقلة)
فريدمان	تقييم الفروق بين مجموعتين أو أكثر	بيانات ترتيبية	• مقاييس متكررة / أزواج متماثلة

الإرشادات العامة لاستخدام SPSS لحساب اختبار: مان وتني U، ولكوكسن T، كروسكال- وليز وفريدمان.

أولا: اختبار مان وتني U:

الإجراء المتبع:

بعد إدخال البيانات Data Entry يقوم الباحث بتحليلها وفقاً للإجراء التالي:

1- في القائمة المعروضة في الجزء العلوي من الشاشة: انقر فوق Analyze ⟸ Non-Parametric tests ⟸ و انقر على Independent 2 Samples.

2- التأكد من أن صندوق Mann-Whitney قد تم تحديده.

3- انقر فوق المتغير التابع، وانقله إلى مربع Test Variable List.

4- انقر فوق المتغير المستقل، وانقله إلى مربع Grouping Variable.

5- انقر فوق زر Define Groups واكتب قيمة Group 1، وقيمة Group 2، وأدخل قيمة 1 و 2 إلى مجموعة الصناديق الملائمة.

6- انقر فوق Contiue.

7- انقر فوق OK.

المخرجات Spss output:

يُنتج هذا الإجراء جدولين يعرض الأول عدد الدرجات، متوسط الرتب ومجموع الرتب لكل مجموعة. أما الجدول الثاني فيقوم بعرض نتائج الاختبار مشتملاً على قيمة U، ودرجة Z التقريبية، وقيمة مستوى الدلالة لكل من مان وتني، و Z (Asymp. sig 2-tailed). لاحظ أنه إذا كان حجم العينة أكبر من 30، فإن برنامج SPSS سيولد قيمة اختبار Z- approximation. فإذا كانت القيمة المحسوبة لـِ U من بيانات العينة أصغر أو مساوية

للقيمة الجدولية، فالباحث إذاً يمكنه الوصول إلى نتيجة مفادها أن الفرق بين المجموعتين ذو دلالة إحصائية.

ثانيا: اختبار ولكوكسن Wilcoxon test:

الإجراء المتبع:

بعد إدخال البيانات Data Entry يقوم الباحث بتحليلها وفقاً للإجراء التالي:

1- في القائمة الموجودة في الجزء العلوي من الشاشة انقر على:
Analyze ⟸ Non-Parametric tests ⟸ وانقر على Related Samples2

2- التأكد من أن صندوق Wilcoxon test قد تم تحديده Test type.

3- انقر فوق المتغيرات التي تمثل الدرجات في كل من 1، 2، وانقلها إلى مربع Test Paired List.

4- انقر فوق OK.

المخرجات Spss output:

إن الأمرين الأساسيين اللَّذَيْن ينبغي على الباحث التركيز عليهما من خلال الجدولين (جدول يعرض العدد "N"، ومتوسط الرتب، ومجموع الرتب - السالبة والموجبة -، والجدول الثاني يعرض نتائج الاختبار مشتملاً على القيمة التقريبية لدرجة Z، ومستوى الدلالة) هما قيمة Z، ومستويات الدلالة المرتبطة بها والتي يعبر عنها بِـ: Asymp. Sig. (2-tailed). فإذا كان مستوى الدلالة أقل أو مساوياً لِـ 05. (05. ، 01. ، 0.001)، فإنه بالتالي يصل إلى قرار مفاده بأن هناك فرقاً ذا دلالة إحصائية.

ثالثا: اختبار كروسكال ـ وليز Kruskal - Wallis:

الإجراء المتبع:

بعد إدخال البيانات Data Entry يقوم الباحث بتحليلها وفقاً للإجراء التالي:

1- في القائمة الموجودة في الجزء العلوي من الشاشة: انقر فوق:

Analyze ⟸ Non-Parametric tests وانقر K Independent Samples .

2- التأكد من أن مربع Kruskal - Wallis قد تم تحديده في جزء Test type.

3- انقر فوق المتغير التابع وانقله إلى مربع Test Variable List.

4- انقر فوق المتغير المستقل، وانقله إلى مربع Grouping Variable.

5- انقر فوق زر Define Range وأدخل القيم Minimum و Maximum في مربع Grouping Variable (على سبيل المثال) إذا استخدمت 1، 2 و 3 لبيان المعالجات الثلاث، أدخل Minimum of 1 و Maximum of 3.

6- انقر فوق Contiue.

7- انقر فوق OK.

المخرجات Spss output:

في هذا الاختبار ينبغي على الباحث التركيز على الجدول الثاني من المخرجات والذي يحتوي على قيمة X^2، ودرجة الحرية، ومستوى الدلالة Asymp.sig. ومن هنا يتوجب على الباحث أن يكون قراره كالتالي:

إذا كان مستوى الدلالة أقل من 0.05، فإن ذلك يعني أن هناك فرقاً دالاً إحصائياً.

رابعا: اختبار Friedman:

الإجراء المتبع:

بعد إدخال البيانات Data Entry يقوم الباحث بتحليلها وفقاً للإجراء التالي:

1- في القائمة الموجودة في الجزء العلوي من الشاشة: انقر فوق

Analyze ⟸ Non-Parametric tests انقر على K Related Samples

2- تأكد من اختبار Friedman test قد تم تحديده في جزء Test type.

3- انقر فوق المتغيرات التي تمثل موضوعات القياس 1، 2، 3.

4- انقر فوق OK.

المخرجات Spss output:

تحتوي المخرجات لهذا الاختبار على جدولين. الجدول الأول يبين متوسط الرتب لكل حالة معالجة. أما الجدول الثاني فيشتمل على نتائج الاختبار: قيمة مربع كاي (X^2)، درجة الحرية (df) ومستوى الدلالة (α) قيمة الفا للاختبار.

أسئلة للمراجعة:

1- رتب الدرجات التالية متى كان ذلك ضرورياً:

14، 3، 4، 0، 3، 5، 14، 3.

2- تجربة استخدمت عينة مؤلفة من N = 25 في العينة الأولى وN = 10 في العينة الثانية. ولكي نجري اختبار مان – وتني U لـ: U = 50 مفترضاً أن هذا العدد هو أصغر قيمة U. ما هي قيمة U العينة الأخرى.

3- في دراسة أوضحت أن درجات عينة الأولاد (4) ودرجات عينة البنات (9) هي كالتالي:

درجة الأولاد: 21، 14، 17، 8

درجة البنات: 18، 25، 23، 21، 34، 28، 32، 30، 13

المطلوب: إجراء اختبار مان – وتني.

4- طبقاً لجدول مان – وتني، فإن قيمة U = 30 ، وهي دالة على مستوى 05. لكلتا العينتين N = 11. وإذا ما تم استخدام هذه القيمة في التوزيع القريب من الطبيعي Normal approximation، هل تولد لنا درجة Z في المنطقة الحرجة؟

5- أراد طبيب أن يختبر فعالية دواء جديد لالتهاب المفاصل وذلك لقياس قوة قبض اليد لدى المرضى قبل وبعد استخدام هذا الدواء. وقد كانت فروق الدرجات لعشرين من المرضى كالتالي:

+ 41 ،−2 ، + 25 ، −8 (لا تغير) 0 ، + 14 + 34 ، + 16 + 46 ، + 3

وتمثل كل درجة الفرق في قوة قبض اليد. وتشير (+) إلى قوة قبض اليد بعد تناول الدواء.

المطلوب: استخدام اختبار T لتحديد ما إذا كانت هذه البيانات تقدم دليلاً كافياً مفاده أن الدواء الجديد لديه تأثير دال.

اختبر هذه البيانات على مستوى $\alpha = 0.05$.

6- البيانات التالية تحتوي على عينة مؤلفة من 7 أشخاص N = 7 وتم اختبار كل فرد من هؤلاء في وضعين مختلفين من المعالجة. والبيانات لهذه العينة يظهرها الجدول التالي:

المشاركون	المعالجة الأولى (1)	المعالجة الثانية (2)	الفرق
1	8	24	+ 16
2	12	10	− 2
3	15	19	+ 4
4	31	52	+ 21
5	26	20	− 6
6	32	40	+ 8
7	19	29	+ 10

المطلوب: إجراء اختبار ولكوكسن من خلال الخطوات التي تعلمتها من مطالعتك لهذا الفصل.

7- البيانات التالية تظهر نتائج جاءت من دراسة استخدمت عينات مستقلة لغرض المقارنة في أوضاع من المعالجات. بعد المعالجة فقد تم ترتيب كل الأفراد في العينات الثلاث. وأن حجم الرتب لِـ: T تم حسابه لكل معالجة.

1	2	3
8	1	11
2	7	5
4	6	9
10	3	12
14	13	15
$T_1 = 38$	$T_2 = 30$	$T_3 = 52$
$n_1 = 5$	$n_2 = 5$	$n_3 = 5$

المطلوب: إجراء اختبار كروسكال – وليز من خلال الخطوات التي تعلمتها من خلال هذا الاختبار؟

8- في دراسة أوضحت نتائجها أن طلاب المدارس الثانوية الذين هم بشكل منتظم يشاهدون بعض الرسوم المتحركة يسجلون درجات عالية مقارنة بنظرائهم الذين لم يشاهدوا هذه الرسوم. وإن نتيجة اختبار T للمقاييس المستقلة، تشير إلى فرق دال بين المجموعتين. ولإعادة البيانات هنا:

أ- هل اختبار مان – وتني، يشير أيضاً إلى فرق دال؟ (لاستخدام مستوى دلالة $\alpha = 0.05$).

ب- عادة ما يجرى اختبار كروسكال – وليز لتقييم الفرق بين ثلاث مجموعات من العينات أو أكثر. ومع هذا فإنه بالإمكان استخدام هذا الاختبار للمقارنة بين مجموعتين من العينات. احسب اختبار كروسكال – وليز مستخدماً درجة دلالة $\alpha = 0.05$ لتقييم الفرق بين هاتين المجموعتين من الطلاب. هل النتيجة التي سنتحصل عليها متسقة مع النتيجة المتحصل عليها من خلال اختبار مان – وتني.

9- مسح اجتماعي لعينة مؤلفة من 7 من الطلاب الجامعيين يطلب منهم أن يرتبوا بشكل منظم الاحتمالات الثلاثة التالية فيما يتعلق بأقرب الأشخاص في علاقاتهم الاجتماعية:

أ- صديق من نفس الجنس.

ب- صديق من الجنس المقابل.

ج- عضو من العائلة من أي من الجنسين.

والبيانات التالية تظهر في الجدول التالي:

الطالب	نفس الجنس	الجنس المقابل	أحد أعضاء العائلة
1	1	3	2
2	1	2	3
3	2	1	3
4	2	3	1
5	2	1	3
6	3	2	1
7	1	2	3
المجموع	12	14	16

المطلوب: إجراء اختبار فريد مان مستخدماً الإجراءات التي تعلمتها من خلال مطالعتك لهذا الفصل (مستوى الدلالة 05. = α).

10- نفترض أننا نرغب في دراسة المقاييس المتكررة لمقارنة ثلاث حالات من المعالجة مستخدماً نفس العينة المؤلفة من N = 6 مشاركين في كل معالجة. يطلب منك التفكير في احتمالين من النتائج المتطرفة لهذه الدراسة.

أ- تخيل أنه لا يوجد تداخل Overlap بين المجموعات الثلاث، وتحديداً. افترض أن كل المشاركين الستة تم ترتيبهم الأول في المعالجة الأولى، والثاني في المعالجة الثانية، والثالث في المعالجة الثالثة، في هذه الحالة، يكون حجم الترتيب للمعالجات الثلاثة هو $R_1=6$, $R_2=12$, $R_3=18$. السؤال المطروح للإجابة عليه هو: ما هي النتيجة المتوقعة لهذه البيانات من خلال اختبار فريد مان. احسب قيمة X^2.

ب- تخيل الآن أنه لا توجد فروق بين المجموعات الثلاثة من المعالجات. على سبيل المثال، فأن المشاركين الثلاث الأول لديهم ترتيب 1، 2 و 3. عبر المعالجات الثلاثة. وأن آخر المشاركين الثلاثة لديهم رتب 3، 2 و 1. في هذه الحالة فإن كل معالجة لديها نفس القيمة R = 12 للمجموع الكلي للرتب.

مرة ثانية: ما هي النتيجة المتوقعة لاختبار فريدمان، ثم أحسب قيمة X^2.

الهوامش والمصادر:

أولا: الهوامش:

1- Frederick J. Gravetter and Larry B. Wallnau , Statistics for The Behavioral Sciencies , 8th ed. , Wadsworth Cengage Learning , 2010 , P. 666.

2- Ibid , PP. 667 - 668.

3- Ibid , PP. 669 - 673.

4- Ibid , P. 672.

5- Ibid , P. 672.

6- Ibid , PP. 675 - 676.

7- Ibid , P. 676.

8- Ibid , P. 676.

9- Ibid , PP. 676 - 677.

10- Ibid , PP. 677 - 678.

11- Ibid , PP. 679 - 680.

12- Ibid , P. 683.

13- Ibid , PP. 684 - 687.

ثانيا: المصادر:

1- Frederick J. Gravetter and Larry B. Wallnau , Statistics for The Behavioral Sciencies , 8th ed. , Wadsworth Cengage Learning , 2010.

2- George Argyrous , Statistics for Social and Health Research With a Guide to SPSS , Sage Publications , London , 2001.

3- George DiEkhoff, Statistics for the social and Behavioral Scienaces: Univariate, Bivariate, Multivariate , MCGRAW-Hill Companies INC. USA , 1992.

4- جولي بالانت (ت) خالد العامري، التحليل الإحصائي باستخدام SPSS، دار الفاروق للنشر والتوزيع، مصر، 2009 م.

الملاحــق

- ملحق (1) الجداول الإحصائية
- ملحق (2) مسرد لبعض المفاهيم الإحصائية الواردة في متن المصنف
- ملحق (3) منظم إحصائي
- ملحق (4) بعض المعادلات الأساسية الواردة في الكتاب

المصادر :

- أولاً : باللغة العربية
- ثانيا : باللغة الإنجليزية

ملحق (1): الجداول الإحصائية

- جدول 1/1 : المساحة تحت المنحنــى الطبيـعي المعياري .
- جدول 2/1 : توزيـعات t .
- جدول 3/1 : توزيـعات F .
- جدول 4/1 : القيـم الحرجة لارتباط بيرسون (r) .
- جدول 5/1 : القيـم الحرجة لارتباط سبيرمان (rs) .
- جدول 6/1 : توزيـعات مربع كـاي (X2) .
- جدول 7/1 : القيـم الحرجة لاختبار مان-وتنــي U .
- جدول 8/1 : القيـم الحرجة لاختبار ولكـوكسن T .

ملحق (2): مسرد المفاهيم الإحصائية الواردة في متن المصنف

ملحق (3) منظم إحصائي.

ملحق (4) بعض المعادلات الأساسية الواردة في الكـتاب.

المصادر

- أولا : باللغة العربية .
- ثانيا : باللغة الإنجليـزيـة .

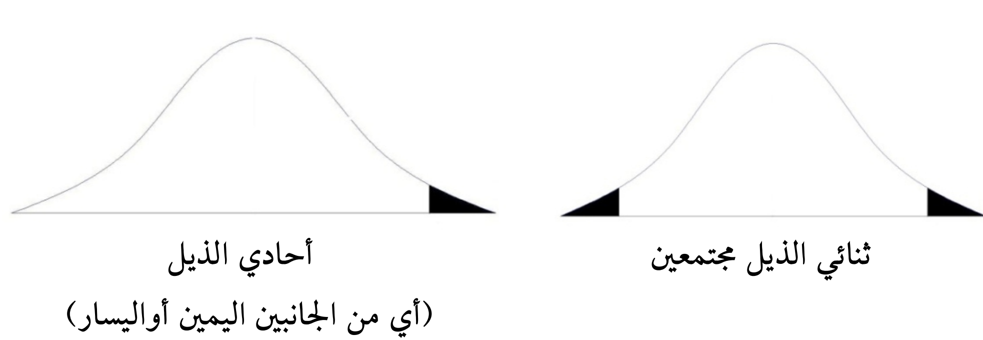
ملحق (1): الجداول الإحصائية

جدول 1/1 المساحة تحت المنحنــى الطبيـعي المعياري

Z	المساحة تحت المنحنى بين كلتا النقطتين	المساحة تحت المنحنى أبعد من النقطتين	المساحة تحت المنحنى أبعد من نقطة واحدة
±0.1	0.080	0.920	0.4600
±0.2	0.159	0.841	0.4205
±0.3	0.236	0.764	0.3820
±0.4	0.311	0.689	0.3445
±0.5	0.383	0.617	0.3085
±0.6	0.451	0.549	0.2745
±0.7	0.516	0.484	0.2420
±0.8	0.576	0.424	0.2120
±0.9	0.632	0.368	0.1840
±1	0.683	0.317	0.1585
±1.1	0.729	0.271	0.1355
±1.2	0.770	0.230	0.1150
±1.3	0.806	0.194	0.0970
±1.4	0.838	0.162	0.0810
±1.5	0.866	0.134	0.0670
±1.6	0.890	0.110	0.0550
±1.645	0.900	0.100	0.0500
±1.7	0.911	0.089	0.0445
±1.8	0.928	0.072	0.0360
±1.9	0.943	0.057	0.0290
±1.96	0.950	0.050	0.0250
±2	0.954	0.046	0.0230
±2.1	0.964	0.036	0.0180

Z	المساحة تحت المنحنى بين كلتا النقطتين	المساحة تحت المنحنى أبعد من النقطتين	المساحة تحت المنحنى أبعد من نقطة واحدة
±2.2	0.972	0.028	0.0140
±2.3	0.979	0.021	0.0105
±2.33	0.980	0.020	0.0100
±2.4	0.984	0.016	0.0080
±2.5	0.988	0.012	0.0060
±2.58	0.990	0.010	0.0050
±2.6	0.991	0.009	0.0045
±2.7	0.993	0.007	0.0035
±2.8	0.995	0.005	0.0025
±2.9	0.996	0.004	0.0020
±3	0.997	0.003	0.0015
±3.1	0.998	0.002	0.0001
±3.2	0.9986	0.0014	0.0007
±3.3	0.9990	0.0010	0.0005
±3.4	0.9993	0.0007	0.0003
±3.5	0.9995	0.0005	0.00025
±3.6	0.9997	0.0003	0.00015
±3.7	0.9998	0.0002	0.00010
±3.8	0.9986	0.00014	0.00007
±3.9	0.99990	0.00010	0.00005
±4	>0.99990	<0.00010	<0.00005





	أحادي الجانب					
	0.25	0.10	0.05	0.25	0.01	0.005
	ثنائي الجانب					
Df	0.50	0.20	0.10	0.05	0.02	0.01
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.499
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.812	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106
12	0.695	1.356	1.782	2.179	2.681	3.055
13	0.694	1.350	1.771	2.160	2.650	3.012
14	0.692	1.345	1.761	2.145	2.624	2.977
15	0.691	1.341	1.753	2.131	2.602	2.947
16	0.690	1.337	1.746	2.120	2.583	2.921
17	0.689	1.333	1.740	2.110	2.567	2.898
18	0.688	1.330	1.734	2.101	2.552	2.878
19	0.688	1.328	1.729	2.093	2.539	2.861
20	0.687	1.325	1.725	2.086	2.528	2.845
21	0.686	1.323	1.721	2.080	2.518	2.831

	أحادي الجانب					
	0.25	0.10	0.05	0.25	0.01	0.005
	ثنائي الجانب					
Df	0.50	0.20	0.10	0.05	0.02	0.01
22	0.686	1.321	1.717	2.074	2.508	2.819
23	0.685	1.319	1.714	2.069	2.500	2.807
24	0.685	1.318	1.711	2.064	2.492	2.797
25	0.684	1.316	1.708	2.060	2.485	2.787
26	0.684	1.315	1.706	2.56	2.479	2.779
27	0.684	1.314	1.703	2.052	2.473	2.771
28	0.683	1.313	1.701	2.048	2.469	2.763
29	0.683	1.311	1.699	0.045	2.462	2.756
30	0.683	1.310	1.697	2.042	2.457	2.750
40	0.681	1.303	1.684	2.021	2.423	2.704
60	0.679	1.296	1.671	2.000	2.390	2.660
120	0.677	1.289	1.658	1.980	2.358	2.617
∞	0.674	1.282	1.645	1.960	2.326	2.576

جدول 3/1 توزيـعات F

مستوى الدلالة 0.05 = α

n_1 / n_2	1	2	3	4	5	6	8	12	24	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	238.9	243.9	249.0	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.41	19.45	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.84	8.74	8.64	8.53
4	7.71	6.94	6.59	6.59	6.39	6.26	6.16	6.04	5.91	5.77
5	6.16	5.79	5.41	5.19	5.05	4.95	4.82	4.68	4.53	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.00	3.84	3.67
7	5.99	4.74	4.35	4.12	3.97	3.87	3.73	3.57	3.41	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.28	3.12	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.07	2.90	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.91	2.74	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.79	2.61	2.40
12	4.75	3.88	3.49	3.26	3.11	3.00	2.85	2.69	2.50	2.30
13	4.67	3.80	3.41	3.18	3.02	2.92	2.77	2.60	2.42	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.53	2.35	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.48	2.29	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.42	2.24	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.38	2.19	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.34	2.15	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.31	2.11	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.28	2.08	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.25	2.05	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.23	2.03	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.38	2.20	2.00	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.18	1.98	1.73
25	4.24	3.38	2.99	2.76	2.60	2.49	2.34	2.16	1.96	1.71
26	4.22	3.37	2.98	2.74	2.59	2.47	2.32	2.15	1.95	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.30	2.13	1.93	1.67
28	4.20	3.34	2.95	2.71	2.56	2.44	2.29	2.12	1.91	1.65
29	4.18	3.33	2.93	2.70	2.54	2.43	2.28	2.10	1.90	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.09	1.89	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.00	1.79	1.51
60	4.00	3.15	2.76	2.52	2.37	2.25	2.10	1.92	1.70	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.02	1.83	1.61	1.25
∞	3.84	2.99	2.60	2.37	2.21	2.09	1.94	1.75	1.52	1.00

جدول 3/1 توزيعات F
مستوى الدلالة 0.01 = α

n_1 / n_2	1	2	3	4	5	6	8	12	24	∞
1	4052	4999	5403	5625	5764	5854	5981	6106	6234	6366
2	98.49	99.01	99.17	99.25	99.30	99.33	99.36	99.42	99.46	99.50
3	34.12	30.81	29.46	28.71	28.24	27.91	27.49	27.05	26.60	26.12
4	21.20	18.00	16.69	15.98	15.52	15.21	14.80	14.37	13.93	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.27	9.89	9.47	9.02
6	13.74	10.92	9.78	9.15	8.75	8.47	8.10	7.72	7.31	6.88
7	12.25	9.55	8.45	7.85	7.46	7.15	6.84	6.47	6.07	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.03	5.67	5.28	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.47	5.11	4.73	4.31
10	10.04	7.56	6.55	5.99	5.54	5.39	5.06	4.71	4.33	3.91
11	9.65	7.20	6.22	5.67	5.32	5.07	4.74	4.40	4.02	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.50	4.16	3.78	3.36
13	9.07	6.70	5.74	5.20	4.86	4.62	4.30	3.96	3.59	3.16
14	8.86	6.51	5.56	5.03	4.69	4.46	4.14	3.80	3.43	3.00
15	9.68	6.36	5.42	4.89	4.56	4.32	4.00	3.67	3.29	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	3.89	3.55	3.18	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.79	3.45	3.08	2.65
18	8.28	6.01	5.09	4.58	4.25	4.01	3.71	3.37	3.00	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.63	3.30	2.92	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.56	3.23	2.86	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.51	3.17	2.80	2.36
22	7.94	5.72	4.82	4.31	3.99	3.76	3.45	3.12	2.75	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.41	3.07	2.70	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.36	3.03	2.66	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.32	2.99	2.62	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.29	2.96	2.58	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.26	2.93	2.55	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.23	2.90	2.52	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.20	2.87	2.49	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.17	2.84	2.47	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	2.99	2.66	2.29	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.82	2.50	2.12	1.60
120	6.85	4.75	3.95	3.48	3.17	2.96	2.66	2.34	1.95	1.38
∞	6.64	4.60	3.78	3.32	3.02	2.80	2.51	2.18	1.79	1.00

جدول 4/1 القيم الحرجة لارتباط بيرسون (r)*

* لكي يكون ارتباط بيرسون(r) ارتباطاً دالاً ، لابد أن تكون قيمة (r) أكبر أو مساوية للقيمة الحرجة في الجدول .

Df=N-2	مستوى الدلالة لاختبار أحادي الجانب			
	.05	.025	.01	.005
	مستوى الدلالة لاختبار ثنائي الجانب			
	.10	.05	.02	.01
1	.988	.997	.9995	.9999
2	.900	.950	.980	.990
3	.805	.878	.934	.999
4	.729	.811	.882	.917
5	.665	.754	.833	.874
6	.622	.707	.789	.834
7	.582	.666	.750	.798
8	.549	.632	.716	.765
9	.521	.602	.685	.735
10	.497	.576	.658	.708
11	.476	.553	.634	.684
12	.458	.532	.612	.661
13	.441	.514	.592	.641
14	.426	.497	.574	.623
15	.412	.482	.558	.606
16	.400	.468	.542	.590
17	.389	.456	.528	.575
18	.378	.444	.516	.561
19	.369	.433	.503	.549
20	.360	.423	.492	.537
21	.352	.413	.482	.526
22	.344	.404	.472	.515
23	.337	.396	.462	.505
24	.330	.388	.453	.496
25	.323	.381	.445	.487
26	.317	.374	.437	.479
27	.311	.367	.430	.471
28	.306	.361	.423	.463
29	.301	.355	.416	.456
30	.296	.349	.409	.449
35	.275	.325	.381	.418
40	.257	.304	.358	.393
45	.243	.288	.338	.372
50	.231	.273	.322	.354
60	.211	.250	.295	.325
70	.195	.232	.274	.302
80	.183	.217	.256	.283
90	.173	.205	.242	.267
100	.164	.195	.230	254

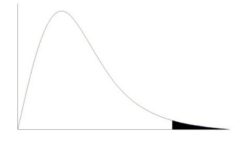
جدول 5/1 القيم الحرجة لارتباط سبيرمان *

*** لكي يكون ارتباط سبيرمان ارتباطاً دالاً ، لابد أن تكون قيمةr_s أكبر أو مساوية للقيمة الحرجة في الجدول .**

	مستوى الدلالة لاختبار أحادي الجانب			
	.05	.025	.01	.005
N	مستوى الدلالة لاختبار ثنائي الجانب			
	.10	.05	.02	.01
4	1.000			
5	0.900	1.000	1.000	
6	0.829	0.886	0.943	1.000
7	0.714	0.786	0.893	0.929
8	0.643	0.738	0.833	0.881
9	0.600	0.700	0.783	0.833
10	0.564	0.648	0.754	0.794
11	0.536	0.618	0.709	0.755
12	0.503	0.587	0.671	0.727
13	0.484	0.560	0.648	0.703
14	0.464	0.538	0.622	0.675
15	0.443	0.521	0.604	0.654
16	0.429	0.503	0.582	0.635
17	0.414	0.485	0.566	0.615
18	0.401	0.472	0.550	0.600
19	0.391	0.460	0.535	0.584
20	0.380	0.447	0.520	0.570
21	0.370	0.435	0.508	0.556
22	0.361	0.425	0.496	0.544
23	0.353	0.415	0.486	0.532
24	0.344	0.406	0.476	0.521
25	0.337	0.398	0.466	0.511
26	0.331	0.390	0.457	0.501
27	0.324	0.382	0.448	0.491
28	0.317	0.375	0.440	0.483
29	0.312	0.368	0.433	0.475
30	0.306	0.362	0.425	0.467
35	0.283	0.335	0.394	0.433
40	0.264	0.313	0.368	0.405
45	0.248	0.294	0.347	0.382
50	0.235	0.279	0.329	0.363
60	0.214	0.255	0.300	0.331
70	0.190	0.235	0.278	0.307
80	0.185	0.220	0.260	0.287
90	0.174	0.207	0.245	0.271
100	0.165	0.197	0.233	0.257

جدول 6/1 توزيـعات مربـع كـاي X^2 *

*** مدخلات الجدول تشير إلى القيم الحرجة لمربع كاي X^2 .**



X^2 الحرجة

	النسبة في المنطقة الحرجة *Proportion in Critical Region*				
Df	0.10	0.05	0.025	0.01	0.005
1	2.71	3.84	5.02	6.63	7.88
2	4.61	5.99	7.38	0.21	10.60
3	6.25	7.81	9.35	11.34	12.84
4	7.78	9.49	11.14	13.28	14.86
5	9.24	11.07	12.83	15.09	16.75
6	10.64	12.59	14.45	16.81	18.55
7	12.02	14.07	16.01	18.48	20.28
8	13.36	15.51	17.53	20.09	21.96
9	14.68	16.92	19.02	21.67	23.59
10	15.99	18.31	20.48	23.21	25.19
11	17.28	19.68	21.92	24.72	26.76
12	18.55	21.03	23.34	26.22	28.30
13	19.81	22.36	24.74	27.69	29.82
14	21.06	23.68	26.12	29.14	31.32
15	22.31	25.00	27.49	30.58	32.80
16	23.54	26.30	28.85	32.00	34.27
17	24.77	27.59	30.19	33.41	35.72
18	25.99	28.87	31.53	34.81	37.16
19	27.20	30.14	32.85	36.19	38.58
20	28.41	31.41	34.17	37.57	40.00
21	29.62	32.67	35.48	38.93	41.40
22	30.81	33.92	36.78	40.29	42.80
23	32.01	35.17	38.08	41.64	44.18
24	33.20	36.42	39.36	42.98	45.56
25	34.38	37.65	40.65	44.31	46.93
26	35.56	38.89	41.92	45.64	48.29
27	36.74	40.11	43.19	46.96	49.64
28	37.92	41.34	44.46	48.28	50.99
29	39.09	42.56	49.59	52.34	52.34
30	40.26	43.77	50.89	53.67	53.67
40	51.81	55.76	63.69	66.77	66.77
50	63.17	67.50	76.15	79.49	79.49
60	74.40	79.08	88.38	91.95	91.95
70	85.53	90.53	100.42	104.22	104.22
80	96.58	101.88	112.33	116.32	116.32
90	107.56	113.34	124.12	128.30	128.30
100	118.50	124.34	129.56	135.81	140.17

جدول 7/1 القيم الحرجة لاختبار مان – وتنــي على مستوى دلالة $\alpha = 0.05$ *

القيم الحرجة لاختبار أحادي الجانب على مستوى α = **0.05** تشير إليها القيم غير المظللة بينما القيم الحرجة لاختبار ثنائي الجانب تشير إليها القيم المظللة . ولكي تكون U دالة يجب أن تكون مساوية أو أقل من القيم الحرجة في الجدول . الشرطات (–) داخل الجدول تشير إلى أنه ليس هناك إمكانية اتخاذ قرار على مستوى دلالة محدد وقيم n_A و n_B .

n_B \ n_A	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0	0
	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
2	–	–	–	–	0	0	0	1	1	1	1	2	2	2	3	3	3	4	4	4
	–	–	–	–	–	–	–	0	0	0	0	1	1	1	1	1	2	2	2	2
3	–	–	0	0	1	2	2	3	3	4	5	5	6	7	7	8	9	9	10	11
	–	–	–	–	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
4	–	–	0	1	2	3	4	5	6	7	8	9	10	11	12	14	15	16	17	18
	–	–	–	0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	13
5	–	0	1	2	4	5	6	8	9	11	12	13	15	16	18	19	20	22	23	25
	–	–	0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
6	–	0	2	3	5	7	8	10	12	14	16	17	19	21	13	15	16	28	30	32
	–	–	1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
7	–	0	2	4	6	8	11	13	15	17	19	21	24	26	28	30	33	35	37	39
	–	–	1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
8	–	1	3	5	8	10	13	15	18	20	23	26	28	31	33	36	39	41	44	47
	–	0	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
9	–	1	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54
	–	0	2	4	7	10	12	15	17	20	23	26	28	31	34	37	39	42	45	48
10	–	1	4	7	11	14	17	20	24	27	31	34	37	41	44	48	51	55	58	62
	–	0	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55
11	–	1	5	8	12	16	19	23	27	31	34	38	42	46	50	54	57	61	65	69
	–	0	3	6	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62
12	–	2	5	9	13	17	21	26	30	34	38	42	47	51	55	60	64	68	72	77
	–	1	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69
13	–	2	6	10	15	19	24	28	33	37	42	47	51	56	61	65	79	75	80	84
	–	1	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76
14	–	2	7	11	16	21	26	31	36	41	46	51	56	60	66	71	77	82	87	92

n_B \ n_A	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	–	1	5	9	13	17	22	26	31	36	40	45	50	55	59	64	67	74	78	83
15	–	3	7	12	18	23	28	33	39	44	50	55	61	66	72	77	83	88	94	100
	–	1	5	10	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90
16	–	3	8	14	19	25	30	36	42	48	54	60	65	71	77	83	89	95	101	107
	–	1	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98
17	–	3	9	15	20	26	33	39	45	51	57	64	70	77	83	89	96	102	109	115
	–	2	6	11	17	22	28	34	39	45	51	57	63	67	75	81	87	93	99	105
18	–	4	9	16	22	28	35	41	48	55	61	68	75	82	88	95	102	109	116	123
	–	2	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112
19	0	4	10	17	23	30	37	44	51	58	65	72	80	87	94	101	109	116	123	130
	–	2	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119
20	0	4	11	18	25	32	39	47	54	62	69	77	84	92	100	107	115	123	130	138
	–	2	8	13	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127

جدول 7/1 القيـم الحرجة لاختبار مان – وتنـي على مستوى ألفا $\alpha = 0.01$ *

القيم الحرجة لاختبار أحادي الجانب على مستوى $\alpha = 0.01$ تشير إليها القيم غير المظللة بينما القيم الحرجة لاختبار ثنائي الجانب تشير إليها القيم المظللة . ولكي تكون U دالة يجب أن تكون مساوية أو أقل من القيم الحرجة في الجدول . الشرطات (–) داخل الجدول تشير إلى أنه ليس هناك إمكانية اتخاذ قرار على مستوى دلالة محدد وقيم nA و nB .

n_B \ n_A	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	–	–	–	–	–	–	–	–	–	–	–	–	0	0	0	0	0	0	1	1
2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0	0
	–	–	–	–	–	–	0	0	1	1	1	2	2	2	3	3	4	4	4	5
3	–	–	–	–	–	–	–	–	0	0	0	1	1	1	2	2	2	2	3	3
	–	–	–	–	0	1	1	2	3	3	4	5	5	6	7	7	8	9	9	10
4	–	–	–	–	–	0	0	1	1	2	2	3	3	4	5	5	6	6	7	8
	–	–	–	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
5	–	–	–	–	0	1	1	2	3	4	5	6	7	7	8	9	10	11	12	13
	–	–	–	1	2	3	4	6	7	8	9	11	12	13	15	16	18	19	20	22
6	–	–	–	0	1	2	3	4	5	6	7	9	10	11	12	13	15	16	17	18
	–	0	0	1	3	4	6	7	9	11	12	14	16	17	19	21	23	24	26	28
7	–	–	–	0	1	3	4	6	7	9	10	12	13	15	16	18	19	21	22	24
	–	–	0	2	4	6	7	9	11	13	15	17	20	22	24	26	28	30	32	34
8	–	–	–	1	2	4	6	7	9	11	13	15	17	18	20	22	24	26	28	30
	–	–	1	3	5	7	9	11	14	16	18	21	23	26	28	31	33	36	38	40
9	–	–	0	1	3	5	7	9	11	13	16	18	20	22	24	27	29	31	33	36
	–	–	1	2	6	8	11	13	16	19	22	24	27	30	33	36	38	41	44	47
10	–	–	0	2	4	6	9	11	13	16	18	21	24	26	29	31	34	37	39	42
	–	–	0	2	4	6	9	11	13	16	18	21	24	26	29	31	34	37	39	53
11	–	–	0	2	5	7	10	13	16	18	21	24	27	30	33	36	39	42	45	48
	–	–	2	5	8	11	14	17	21	24	28	31	35	38	42	46	49	53	56	60
12	–	–	1	3	6	9	12	15	18	21	24	27	31	34	37	41	44	47	51	54
	–	0	2	5	9	12	16	20	23	27	31	35	39	43	47	51	55	59	63	67
13	–	–	1	3	7	10	13	17	20	24	27	31	34	38	42	45	49	53	56	60

n_B \ n_A	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	–	0	2	6	10	13	17	22	26	30	34	38	43	47	51	56	60	65	69	73
14	–	–	1	4	7	11	15	18	22	26	30	34	38	42	46	50	54	58	63	67
	–	0	3	7	11	15	19	24	28	33	37	42	47	51	56	61	66	70	75	80
15	–	–	2	5	8	12	16	20	24	29	33	37	42	46	51	55	60	64	69	73
	–	0	3	7	12	16	21	26	31	36	41	46	51	56	61	66	71	76	82	87
16	–	–	2	5	9	13	18	22	27	31	36	41	45	50	55	60	65	70	74	79
	–	0	4	8	13	18	23	28	33	38	44	49	55	60	66	71	77	82	88	93
17	–	–	2	6	10	15	19	24	29	34	39	44	49	54	60	65	70	75	81	86
	–	0	4	9	14	19	24	30	36	41	47	53	59	65	70	76	82	88	94	100
18	–	–	2	6	11	16	21	26	31	37	42	47	53	58	64	70	75	81	87	92
	–	1	4	9	15	20	26	32	38	44	50	56	63	69	75	82	88	94	101	107
19	–	0	3	7	12	17	22	28	33	39	45	51	56	63	69	74	81	87	93	99
	–	1	5	10	16	22	28	34	40	47	53	60	67	73	80	87	93	100	107	114
20	–	0	3	8	13	18	24	30	36	42	48	54	60	67	73	79	86	92	99	105

جدول 8/1 القيم الحرجة لاختبار ولكوكسن * Wilcoxon T

* لكي يكون الاختبار اختباراً دالاً يجب أن تكون درجة الاختبار مساوية أو أقل من القيمة الحرجة . (-) تعني هذه الإشارة في العمود أنه لا يوجد قرار ممكن لدلالة α و N .

	مستوى الدلالة لاختبار أحادي الجانب					مستوى الدلالة لاختبار أحادي الجانب			
	.05	.025	.01	.005		.05	.025	.01	.005
	مستوى الدلالة لاختبار ثنائي الجانب					مستوى الدلالة لاختبار ثنائي الجانب			
n	.10	.05	.02	.01	n	.10	.05	.02	.01
5	0	-	-	-	28	130	116	101	91
6	2	0	-	-	29	140	126	110	100
7	3	2	0	-	30	151	137	120	109
8	5	3	1	0	31	163	147	130	118
9	8	5	3	1	32	175	159	140	128
10	10	8	5	3	33	187	170	151	138
11	13	10	7	5	34	200	182	162	148
12	17	13	9	7	35	213	195	173	159
13	21	17	12	9	36	227	208	185	171
14	25	21	15	12	37	241	221	198	182
15	30	25	19	15	38	256	235	211	194
16	35	29	23	19	39	271	249	224	207
17	41	34	27	23	40	286	264	238	220
18	47	40	32	27	41	402	279	252	233
19	53	46	37	32	42	319	294	266	247
20	60	52	43	37	43	336	310	281	261
21	67	58	49	42	44	353	327	296	276
22	75	65	55	48	45	371	343	312	291
23	83	73	62	54	46	389	361	328	307
24	91	81	69	61	47	407	378	345	322
25	100	89	76	68	48	426	396	362	339
26	110	98	84	75	49	446	415	379	355
27	119	107	92	83	50	466	434	397	373

ملحق (2)

مسرد لبعض المفاهيم الواردة في متن المصنف

تسهيلاً على طلابنا، فقد ارتأينا أن نضمن في هذا الكتاب بعض المفاهيم الإحصائية مع تعريفاتها:

1- ***Arithmetic Mean*** وسط حسابي: الوسط الحسابي هو مجموع كل الدرجات في التوزيع مقسماً على العدد الكلي للحالات.

2- ***Asymmetric Measures of association*** مقاييس التطابق اللا متماثلة: هي تلك المقاييس التي تعتمد قيمتها على أي من المتغيرات الذي حدد بأنه متغيرٌ مستقلٌ، وأي من المتغيرات الذي حدد بأن يكون تابعاً.

3- ***Binomial distribation*** التوزيع ذو الحدين: هو ذلك التوزيع الذي يمتلك فقط قيمتين أو فئتين احتماليتين.

4- ***Bivariate descriptive Statistics*** الإحصاءات الوصفية الثنائية: هي نوع من الإحصاءات التي يمكن استخدامها في التحليل لمعرفة ما إذا كانت هناك علاقة بين متغيرين.

5- ***Bivariate table*** الجدول الثنائي: هو ذلك الجدول الذي يعرض التوزيع التكراري المشترك لمتغيرين.

6- ***Case*** حالة: هي كينونة تعرض أو تمتلك خاصيات متغير. أي خاصية تعكس عملية القياس المرتبطة بالمتغير.

7- ***Central Limit Theorm*** نظرية النهائية المركزية: هي تلك النظرية التي تبين أنه إذا كان عدد العينات العشوائية المأخوذة من مجتمع ما، وذات أحجام متساوية واللا متناهية العدد، فإن توزيع المعاينة لمتوسطات العينات يميل إلى أن يكون موزعاً – تقريباً – توزيعاً طبيعياً.

8- **Class Interval** فئات متساوية: الفئات المتساوية هي عبـارة عـن مـدى مـن القـيم على توزيع تم تصنيف هذه القيم معاً للعرض والتحليل.

9- **Coefficient of relative Variation (CRV)** معامل التباين النسبي: معامل التباين النسبي هو عبارة عن إحـصاءات وصـفية تعـبر عـن الانــحراف المعيـاري للتوزيـع كنسبة للمتوسط الحسابي.

10- **Determination (r^2)** معامل التحديد (r^2): هـي نـسبة كـل التبـاين في y الـذي تم تفسيره من خلال X. ويمكننا إيجاد هذا التباين بتربيع قيمة ارتباط بيرسون (r).

11- **Coefficient of Mnltiple determination(R^2)** معامل التحديد المتعدد: هـو ذاك الإحصاء المساوي للتباين الكلي الذي تم تفـسيره في المـتغير التـابع مـن خـلال كـل المتغيرات المستقلة مجتمعة.

12- **Colum** عمود: العمـود هـو البعـد العمـودي للجـدول الثنـائي. وحـسب العـرف الأكاديمي يمثل كل عمود درجة على المتغير المستقل.

13- **Conceptual definitions** التعريفـات التـصورية: التعريفـات التـصورية هـي تلـك التعريفات التي تحدد خاصيات متغير (يطلق عليها أيضاً التعريف الاسمي).

14- **Concordant Pair** الأزواج المتوافقة: تعني الأزواج المتوافقـة أن حـالتين في توزيـع مشترك تم ترتيبهما على كلا المتغيرين بشكل متساوٍ.

15- **Confidence Level** مستوى الثقة: يعني مستوى الثقة الاحتماليـة بـأن تقـدير فـترة سوف يتضمن قيمة معلمة المجتمع التي تم تقديرها.

16- **Constant** ثابت: الثابت هو عبارة عن الخاصية التي لا تتبـاين مـن حالـة إلى حالـة أخرى.

17- **Contingency Table** جدول التوافق: (انظر تعريف الجدول الثنائي في هذا المسرد).

18- **Continous Variable** المتغير المتصل: المتغير المتصل هو ذلـك المـتغير الـذي يأخـذ جميع القيم بين حدي المتغير، وإذا زاد من رقم إلى الرقم التالي له يمـر بكـل الكـسور الممكنة بينهما (العمر على سبيل المثال).

-19 ***Coordinate*** متساوٍ: هي نقطة على شكل الانتشار تـشير بـشكل تلقـائي إلى القـيم لحالة معينة تدون لكل متغير.

-20 ***Correlation Matrix*** مصفوفة الارتباط: عبارة عن جدول يـبين الارتباطـات بـين كل المتغيرات الممكنة المؤتلفة.

-21 ***Critical Region*** المنطقة الحرجة: تعني المنطقة الحرجة مدى الـدرجات الـذي تَـسَبَّبَ في رفض الفرض الصفري (Ho) على مستوى محدد من الدلالة (α).

-22 ***Crosstabulation*** التقاطع: (انظر تعريف الجدول الثنائي في هذا المسرد).

-23 ***Cumulative Frequency Table*** الجدول التكراري التجمعي (التراكمـي): يـشير الجدول التكراري التجمعي إلى كل قيمة في التوزيع وتجمع أعداد الفئات إلى بعـضها البعض لتشمل تلك القيمة.

-24 ***Cumulative relative Frequency Table*** الجدول التكراري المجمع النسبي: هي تلك الجداول التي تعبر عن الحالات داخـل كـل قيمـة مـن قـيم المـتغير كنـسبة مئويـة أو تناسب من المجموع الكلي للحالات تضاف إلى بعضها البعض لتشمل تلك القيمة..

-25 ***Data*** بيانات: هي تلك المعلومات التي يقوم الباحث بجمعهـا مـن خـلال قـضية أو مشكلة يطرحها للبحث والاستقصاء.

-26 ***Degree of Freedom*** درجـة الحريـة: مفهـوم إحـصائي يـشير إلى عـدد القـيم في التوزيـع الـذي يكـون حـراً في تباينـه فيمـا يتعلـق بمتوسـط عينـة، توزيـع لـه N-1 درجة حرية، هذا يعني أنه فيما يتعلـق بقيمـة محـددة المتوسـط. N-1 درجـات حرة في تباينها. على سبيل المثال، لو كان لدينا متوسط يساوي 3 و N= 5. فالتوزيع لخمس درجات سيكون لديها N-1 أو 4 درجات حرية. عندما تكون قيم الدرجات 1، 2، 3، 4، فإن الخامسة يجب أن تكون 5 ولا قيم أخرى.

-27 ***Dependent Sample*** عينات تابعة: العينات التابعـة هـي تلـك العينـات الـتي يـتم تحديدها من خلال تركيبة عينات أخرى.

28- *Descriptive Statistics* إحــصاءات وصــفية: تــشير الإحــصاءات الوصــفية إلى الأعداد والتمثيل البياني، وتقنيات الجدولة لتنظيم البيانات، وعرضها، وتحليلها.

29- *Dichotomous Variable* المتغير الثنــائي: هــو ذلــك المــتغير الــذي يمتلــك قيمــتين احتماليتين فقط.

30- *Directional Hypothesis* الفرضية الاتجاهية: هي تلك الفرضية التي يتوقع الباحـث من خلالها وجود علاقة محددة بين متغيرين (سالبة أو موجبة).

31- *Disconcordant Pair* الأزواج المتنافرة: تعني الأزواج المتنافرة أن حالتين في توزيع مشترك تم ترتيبهما على متغير بشكل مختلف عن بعضهما البعض.

32- *Discrete Variable* المتغير المنفصل: هو ذلك المتغير الذي لا تُمَكِّنُهُ طبيعته مـن أن يأخذ جميع القيم بـين حـدي المـتغير، بـل تتحـرك عـن أعـداد معينـة دون سـواها ؛ وبالتالي فإن المتغير المنفصل لا يحتوي على كسور (عـدد الأطفـال في الأسـرة علـى سبيل المثال).

33- *Dummy Variable* المتغيرات الديميوية: هي متغيرات مقاسة على أي مـستوى مـن القياس بما فيها المقياس الاسمي. ولديها فئتان يرمز للفئة الأولى بـصفر (0)، والفئـة الثانية بِـ (1) وذلك لتضمينها في معادلة الانــحدار.

34- *Effect Size* حجم التأثير: عبارة عن مجموعـة مـن الإحـصاءات الـتي تحـدد القـوة النسبية للفروق بين قيم الوسط الحسابي. بمعنى آخر، يعني حجم التأثير مقدار التباين الكلي في المتغير التابع الذي يمكن التنبؤ به من معرفة مستويات المتغير المستقل.

35- $Eta^2 = n^2$ إيتا تربيع: يمثل إحصاء Eta^2 نسبة التباين في المـتغير التـابع الـتي يفـسرها ويوضحها المتغير المستقل، وتتراوح قيمتها من 0 إلى 1.0.

36- *Frequency* التكرار: هو عدد المرات التي تظهر فيها درجة مفردة في مجموعة بيانات.

37- *Frequency Table* الجدول التكراري: هي تلـك الجـداول الـتي تلخـص إجابـات المبحوثين في كل فئة من فئات المتغير المدروس.

38- *Gamma* جاما (G): تعتبر جاما مقياساً للتطـابق للمـتغيرات الترتيبيـة (والثنائيـة). ويشار إلى جاما (G) في بعض الأحيان بـِ Goodman-Kruskalś Gamma كمقيـاس متماثل للتطابق. وتتراوح قيمة جاما من 1 − إلى 1.0+. وتشير قيمة 1+ إلى تطـابق تـام، حيث إن كل التباين في المتغير التابع يعزى إلى التباين في المتغير المستقل. أما قيمة 1− فهي تشير، مرة أخرى، إلى وجود تطابق حتمي، ولكنه تطـابق سـالب. بينمـا كـل التبـاين في المتغير التابع يعزى للتباين في المتغير المستقل، فالتطابق يتجه في الاتجاه المعاكس (سالب). وعندما تشير قيمة جاما إلى صـفر (0) يعـني لا يوجـد تطـابق، لا شـيء في التبـاين في المتغير التابع يمكن أن يفسر من خلال التباين في المتغير المستقل.

39- *Goodness of fit test* اختبار جودة التطـابق: يبحـث هـذا الاختبـار مـا إذا كانـت بيانات عينة ما (fo) تتطابق مع توقعات معينة (fe) يُفترض أن يتميز بها مجتمع ما.

40- *Hypothesis* الفرض: هو عبارة عن بيان حول بعض خصائص التوزيـع في مجتمـع إحصائي.

41- *Hypothesis testing* اختبار الفرض: هو الإجـراء الـذي يحـدد مـن خلالـه مـا إذا كانت بعض مظاهر توزيع مجتمع إحصائي تمتلك خاصيات محددة.

42- *Independence* الاسـتقلال: يعتـبر المـتغيران مـستقلين إذا كـان نمـط التبـاين في الدرجات غير مرتبط بنمط التباين في الدرجات للمتغير الآخر.

43- *Index of qualitative Variation (IQV)* مؤشـر التبـاين النسبي: هـو نـوع مـن الإحصاء يخبرنا كم من المتغيرّية توجد في متغير لديه فئات اسمية وأن مدى هذا المقيـاس (IQV) تمتد قيمته من 0 – 1.0. وكلما كانت القيمة قريبة من صـفر، كانـت هنـاك أقـل متغيرية في توزيع المبحوثين على المتغير، وكلما قربت القيمة من 1.0، زادت المتغيرّية.

44- *Inferential Statistics* الإحـصاء الاسـتنتاجي (الاسـتدلالي): يـشير الإحـصاء الاستنتاجي (الاستدلالي) إلى التقنيات العدديـة الـتي تـستخدم للوصـول إلى نتـائج حول توزيع مجتمع إحصائي استناداً على بيانات من عينة عـشوائية، تم سـحبها مـن ذلك المجتمع الإحصائي.

45- ***Interqnartile range*** المدى الربيعي: هو الفرق بين الحـدود العليـا للربيـع الأول، والربيع الثالث ؛ أي المدى الوسط 50 % للحالات في مجموعة مرتبـة مـن التوزيـع ($R = Q_3 - Q_1$).

46- ***Interval Scale*** المقياس ذو المسافات المتساوية: هو المقياس الـذي يمتلـك وحـدات تم قياسها على ذي المسافات المتساوية ولديها مسافة متساوية بين القيم في المقياس.

47- ***Least Significance difference (LSD)*** أقل فرق الدلالة: هي قيمة عددية يمكن حسابها إذا كان التكرار داخل كل معاملة متساوياً:

$$LDS = t\sqrt{\frac{2\,s^2}{r}}$$

أما إذا كانت التكرارات داخل كل معاملة غير متساوية:

$$LDS = t\sqrt{\frac{S^2}{r_1} + \frac{S^2}{r_1}}$$

48- ***Lambda*** لامبيــدا: هــي مقيــاس للتطــابق لنــسبة التخفــيض في الخطــأ (PRE) للمتغيرات الاسمية. وتتراوح قيمتها من 0 إلى 1.0. وعندما تكـون لامبيـدا مـساوية لصفر (0) يعني أنه لا يوجد تطابق ؛ أي لا شيء في التباين في المتغير التابع يمكـن أن يفسر من خلال التباين في المتغير المستقل. أما إذا كانت قيمة لامبيدا مساوية لِـ 1.0، فهذا يعني وجود تطابق تام 100 % أي أن كل التباين في المتغير التابع يمكـن تفـسيره بواسطة التباين في المتغير المستقل.

49- ***Least Square Method*** طريقة أقل التربيعات: طريقة تـستخدم في توضـيح الخـط المستقيم الأمثل الذي يمثل العلاقة بين المتغير X والمتغير y.

50- ***Level of Significance*** مستوى الدلالة: هو أقصى احتمال الذي يحتمل بمقتـضاه مخاطرة الوقوع في الخطأ من النوع الأول. أي القيمة المحددة التي علـى ضـوئها يُقبـل أو يُرفض الفرض الصفري (Ho).

51- ***Measurment*** القياس: هـو العمليـة الـتي مـن خلالهـا يـستطيع الباحـث أن يحـدد ويسجل الخاصيات الممكنة للمتغير كحالة مفـردة. بمعنـى آخـر، يعـني القيـاس بأنـه طريقة خاصة تتبع في قيـاس المـتغيرات والمفـاهيم الاجتماعيـة (مـتغير النـوع يأخـذ قيمتين ذكراً أو أنثى، ويحدد لنا القياس ما هي الفئة التي يقع فيها هذا الشخص).

52- ***Measurment of association*** مقياس التطابق: إحصاءات وصفية تشير إلى مـدى التغير في قيم متغير مرتبطة بالتغير في قيم متغير آخر.

53- ***Measures of Central Tendency*** مقـاييس النـــزعة المركزيـة: هـي تلـك الإحصاءات الوصفية التي تشير إلى قيمة المعدل في التوزيع.

54- ***Measures of dispersion*** مقاييس التشتت: هي تلك الإحصاءات الوصـفية الـتي تشير إلى تباين الدرجات في التوزيع.

55- ***Median*** الوسيط: هـو أحـد مقـاييس النـــزعة المركزيـة الـذي يـشير إلى القيمـة في مجموعة مرتبة، وتقسم التوزيع إلى نصفين.

56- ***Mode*** المنوال: هو القيمة الأعلى تكراراً في التوزيع. أي القيمة التي تتكرر أكثر مـن غيرها.

57- ***Multiple Correlation*** الارتباط المتعدد: عبارة عن تقنيات إحصائية تشير إلى قـوة بين متغير تابع واثنين أو أكثر من المتغيرات المستقلة.

58- ***Multivariate Regression*** الانـحدار المتعدد: هو تلـك التقنيـات الـتي تبحـث في العلاقة بين متغيرين مستقلين أو أكثر وعلاقتها بمتغير تابع واحد.

59- ***Nominal definition*** التعريف الاسمي: (انظر تعريف التعريف التصوري في هـذا المسرد).

60- ***Nominal Scale*** المقياس الاسمي: مجموعة فئـات لـديها أسمـاء مختلفـة والقيـاس على المستوى الاسمي يصف ويصنف المشاهدات، ولكنه لا يحدد أي تمييز كمي بين هذه المشاهدات.

61- ***Non-directional Hypothesis*** الفرضية غير الاتجاهية: هي تلك الفرضية الـتي لا يتوقع الباحث من خلالها أي علاقة محددة بين متغيرين خاضعين للدراسة.

62- ***Nonparametric test*** الاختبار اللا معلمي (اللا بـارامتري): هـو اختبـار لغـرض حول خصائص توزيع مجتمع أكثر منه كاختبار معلمات مجتمع.

63- ***Normal distribution*** التوزيع الطبيعي: هو التوزيع النظري للـدرجات المتماثلـة، أحادية النموذج، وذات الشكل الجرسي، والمنحنى الطبيعـي المعيـاري دائمـاً يكـون متوسطه صفراً، وانـحرافه المعياري واحد (1).

64- ***Operational definition*** التعريف الإجرائي: هو مجموعة الإجراءات الـتي تـصف النشاطات التي يجب القيام بها من أجل تعيين الأبعاد التي يمكن قياسـها وملاحظتهـا من أجل التعرف على مـا يـشير إليـه هـذا المفهـوم أو ذاك في الدراسـة الـتي يجريهـا الباحث.

65- ***Ordinal Scale*** مقياس تـرتيبي: هـو أحـد مـستويات القيـاس ويـشير، إضـافة إلى خاصية التصنيف، أنه يسمح للحالات أن ترتب وفقاً لدرجة كل فئة، وطبقاً لقيـاس المتغير محل اهتمام الباحث.

66- ***Ordinary Least Square Regression***: عبارة عن قاعدة تبين أفضل خط مطـابق للانـحدار الخطي، ذلك الخط الذي يقلل مجموع تربيعات المتبقي.

67- ***Parameter*** معلمة: المعلمة عبارة عن إحصاء يصف بعض معالم المجتمع: المتوسـط، الانـحراف، التباين... الخ. (متوسط الدخل، والتوزيع العمري لمدينة ما).

68- ***Parametric Test*** اختبار مَعْلمي (بـارامتري): الاختبـار المعلمـي هـو عبـارة عـن اختبار لفرض حول معلمات توزيع مجتمع إحصائي.

69- ***Partial Correlation*** الارتباط الجزئي: هو تلك المعامل التي يستطيع الباحث مـن خلالها تحييد تأثير المتغير الثالث عندما يبحث في العلاقة بين متغيرين X و y.

70- ***Partial Correlation Coefficient*** معامل الارتباط الجزئي: معامل الارتباط الجزئـي هو إحصاء يبين العلاقة بين مـتغيرين، بينمـا الـتحكم في مـتغيرات أخـرى. والرمـز $r_{yx.A}$ يشير لمعامل الارتباط الجزئي عند التحكم في متغير واحد.

71- ***Percentage*** النسب: النسبة عبارة عن إحصاء يعاير الحجم الكلي للحـالات علـى أساس قيمة مائة (100).

72- ***Perfect Association*** تطابق تام: التطابق التام يعبر عن علاقة إحـصائية حيـث إن كل الحالات التي لديها قيم مفردة لمتغير واحد يمتلك قيمة محددة للمتغير الآخر.

73- ***Population*** مجتمع إحصائي: هو مجموع الحالات التي يرغب الباحث في دراستها.

74- ***Post-hoc Tests***: هي اختبارات إضافية التي تجرى بعد اختبـار ANOVA لتحديـد أي من فروق المتوسط تكون دالة وأي منها لا تكون كذلك.

75- ***Proportional Reduction in Error*** نسبة التخفيض في الخطـأ: نـسبة التخفـيض في الخطأ هو المنطق الذي يشكل التعريـف وحـساب لامبيـدا وجامـا. وهـو الإحـصاء الذي يقارن عدد الأخطاء التي حدثت عند التنبؤ بالمتغير التـابع عنـد تجاهـل المـتغير المستقل، مع عدد الأخطاء التي حدثت عند الأخذ في الاعتبار المتغير المستقل.

76- ***Proportions*** التناسب: عبارة عن إحصاء لمعـايرة الحجـم الكلـي للحـالات علـى أساس قيمة واحد (1).

77- ***Random Selection*** اختبار عشوائي: هي عملية المعاينـة حيـث إن كـل مفـردة في المجتمع يكون لديها نفس فرصة الاختيار في العينة.

78- ***Range*** المدى: هو الفرق بين أكبر قيمة وأصغر قيمة في التوزيع.

79- ***Rank*** ترتيب: هو العدد الذي يشير إلى وضع حالة في سلسلة مرتبة.

80- ***Ratio Scale*** المقياس النسبي: هو أحد مستويات القياس الـذي يعطـي قيمـة صـفر (0) للحالات التي تظهر لا كمية في المتغير.

-81 ***Region of Rejection*** منطقة الرفض (انظر Critical region من هذا المسرد).

-82 ***Regression Coefficient*** معامــل الانــحدار: معامــل الانــحدار عبــارة عــن إحصاءات وصفية تشير إلى كم من الوحدات في المتغير التابع ستتغير مع تغير وحـدة معطاة في المتغير التابع.

-83 ***Regression Line*** خط الانـحدار: هو أفضل خط تطابق مستقيم يلخـص العلاقـة بين متغيرين. فخطوط الانـحدار مطابقة نقاط البيانات عـبر مقيـاس أقـل المربعـات الذي به يلمس فيه الخط كل متوسط شرطي لِـ: y. أو تكون قريبة من الخط إلى حـد ممكن لكل الدرجات. (ويصلح خط الانـحدار في عمليـة التنبـؤ مـن خـلال معرفـة قيمة أحد المتغيرين، فإنه بالتالي يمكن التنبؤ بقيم المتغير الثـاني. ويمكـن رسـم خـط الانـحدار عن طريق معامل الانـحدار).

-84 ***Relative Frequencies*** التوزيعات النسبية: هي إحصاءات تبين عدد الحـالات داخـل كل قيمة متعلقة بمتغير ما كنسبة أو تناسب من العدد الكلي للحالات.

-85 ***Research Hypothesis*** الفـرض البحثـي: الفـرض البحثـي عبـارة عـن صـياغة الفرض الصفري (Ho) وفي سياق دلالة الاختبـارات لعينـة مفـردة، يـشير الفـرض البحثي إلى أن المجتمع الذي سحبت منه العينة لا يمتلك خاصية أو قيمة محددة.

-86 ***Residual*** متبقٍ: المتبقي هو الفرق بين القيمة المـشاهدة والقيمـة المتوقعـة لمـتغير مـا خاضع للدراسة.

-87 ***Sample*** عينة: العينة هي مجموعة الحـالات الـتي لا تحتـوي علـى كـل الحـالات في المجتمع الإحصائي، وإنما جزء من هذا المجتمع.

-88 ***Sampling Distribution*** توزيع المعاينة: هـي توزيعـات تكراريـة يمكـن الحـصول عليها من حساب المتوسطات لكل العينات النظريـة المحتملـة للحجـم المعـين الـذي أمكن سحبه من مجتمع إحصائي محدد.

89- **Sampling Error** خطأ المعاينة: يعني خطأ المعاينة درجة الخطأ المتوقعة في احتماليـة المعاينة لعينة محددة التصميم. والمعادلة التي تحـدد خطـأ المعاينـة تحتـوي علـى ثلاثـة عوامل: المعلمة، حجم العينة، والخطأ المعياري.

90- **Scatter Plot** رسم انتشاري: هو عبارة عن تقنية بيانية لوصف توزيع مشترك لعدد متغيرين متصلين.

91- **Somer's d**: هو عبارة مقيـاس توزيعـات غـير متماثلـة التطبيـق يأخـذ في الحـسبان الرتب المشتركة علـى المـتغير y دون المـتغير X إلى جانـب الحـسابات المـستخدمة في معامل GAMMA. يضاف T_y إلى مقام معادلة جاما. وتشير T_y إلى الرتب المتعادلـة على المتغير y دون المتغير X.

ولما كان مقياس سومرز d يأخـذ في اعتبـاره الرتـب المتعادلـة علـى y دون X، فقـد حدث تناقض شديد في درجة الارتباط عند مقارنته بجاما.

92- **Spurious Relationship** علاقة كاذبة: العلاقـة الكاذبـة هـي علاقـة متعـددة بـين متغيرات حيث لا توجد بينها علاقة سببية حقيقية بين المتغيرات المستقلة والمـتغيرات التابعة. فالعلاقة بين هذه المتغيرات كانت نتيجة لبعض المتغيرات الأخرى.

93- **Standard Deviation** انـحراف معياري: الانـحراف المعيـاري هـو أحـد مقـاييس التشتت الذي يشير إلى الجذر التربيعي للتباين.

94- **Standard Error** الخطـأ المعيـاري: الخطـأ المعيـاري هـو الجـذر التربيعـي لمتوسـط مجموع مربعات انـحراف القيم عن متوسطها الحسابي.

95- **Standardizing Distribution** معايرة التوزيع: تعني معايرة التوزيع ذلك التوزيـع الطبيعي الذي يتم فيه تحويل الدرجات الإحـصائية إلى درجـات معياريـة تعـبر عـن درجـات أي توزيـع تكـراري طبيعـي. وتـسمح إجـراءات المعـايرة بقيـاس كـل التوزيعات الطبيعيـة في إطـار الوحـدات الـشائعة، وحـدات الانــحراف المعيـاري، بغض النظر عن الوحدات التي يتم قياسها.

96- **Statistics** الإحصاء: عبارة عن تقنيات رياضية تستخدم لفحص البيانات من أجـل الإجابة على السؤال المطروح أو اختبار نظريات.

97- **Symmetric Measures of association** مقاييس التطابق المتماثل: مقاييس التطابق المتماثل هي تلك المقـاييس الـتي تكـون قوتهـا متـساوية بغـض النظـر عـن أي مـن المتغيرين قد حدد كمتغير مستقل أو أي منهما قد حدد كمتغير تابع.

98- **Tabulation** جدولة: الجدولة هي العملية التي بمقتضاها يستطيع الباحث وضع كل حالة من الحالات المدروسة في المكان المخصص لها في التوزيع التصنيفي.

99- **Test of Independence** اختبار الاستقلال: اختبار الاستقلال هو دراسـة العلاقـة أو العلاقات بين ظاهرتين أو أكثر، وبيان ما إذا كان توزيع هـاتين الظـاهرتين أنهمـا مستقلتان أو غير مستقلتين عن بعضهما البعض.

100- **Test Statistic** إحصائي الاختبار: إحصائي الاختبار هو قيمة إحصائية تسحب من بيانات العينة المدروسة، وتستخدم كأداة للوصول إلى قـرار قبـول أو رفـض فرضـية إحصائية معطاة، وذلك اعتمـاداً علـى القيمـة الإحـصائية المـسحوبة. وتأخـذ هـذه القيمة الإحصائية عديداً من القيم الاحتمالية اعتماداً علـى طبيعـة العينـة العـشوائية المستخدمة في الدراسة.

101- **Total Variation** التباين الكلي: هو انتشار درجات y حول متوسط y.

102- **Type I Error or (Alph Error)** الخطأ من النوع الأول (خطأ ألفـا): يعـني الخطـأ من النوع الأول رفض الفرض الصفري "لا فـرق" عنـدما يكـون في الواقـع فرضـاً صحيحاً.

103- **Type II Error or (Beta Error)** الخطأ من النوع الثاني أو خطأ بيتا: يعـني الخطـأ من النوع الثاني أن الباحث ليس في مقدوره رفض الفرض الـصفري عنـدما يكـون في الواقع فرضاً خطأ.

104- **_Variable_** متغير: المتغير هو عبارة عن خاصية تجريبية تتخذ قيمتين أو أكثر. فإذا كانت هذه الخاصية قابلة للتغير كماً أو نوعاً ينظر إليها كمتغير (العمر، الدخل).

105- **_Variability_** المتغيّرية: المتغيرية تمدنا بقياس كمي لمعرفة إلى أي مدى تكون الدرجات في التوزيع قد انتشرت أو تجمعت معاً، ومن هنا تعرف المتغيّرية من خلال المسافة، حيث يمكننا معرفة كم المسافة التي نتوقعها بين درجة وأخرى. وكم المسافة المتوقعة بين درجة الفرد والمتوسط الحسابي. ويعتبر مفهوم المتغيّرية مفهوماً مهماً في الإحصاءات الاستدلالية. وتحتوي المتغيرية على المدى، المدى الربيعي، والانحراف المعياري. بمعنى آخر، إذا كانت الدرجات في التوزيع كلها متساوية حينئذٍ لا توجد متغيّرية. أما إذا تمت ملاحظة فروق بسيطة بين الدرجات عندئذٍ تكون المتغيّرية قليلة. وأما إذا تمت ملاحظة فروق كبيرة بين الدرجات حينها تكون المتغيّرية كبيرة.

106- **_Variance_** تباين: التباين هو الإحصاء الذي يعبر عن متوسط انحراف الدرجات عن المتوسط الحسابي للتوزيع.

107- **_y - Intercept(a)_** : هي النقطة التي يمر بها خط الانحدار لمحور y.

108- **_Z - Score_** درجة *Z*: درجة Z هي الدرجات المعيارية التي تعبر عن المسافة بين نقطة. ومتوسط توزيع طبيعي كنسبة للانحراف المعياري من ذاك التوزيع الطبيعي والنظري.

ملحق (3)
منظم إحصائي

الملحق التالي يبين لنا ملخصاً منظماً للأساليب الإحصائية الـتي تم تغطيتهـا في هـذا الكتاب. وقد قُسم هذا المنظّم الإحصائي إلى أربعة أقسام. وكـل قـسم مـن هـذه الأقـسام يوضح الأساليب الإحصائية التي تخدم غاية مشتركة. والأقسام الأربعة هي

1- الإحصاء الوصفي.
2- الاختبارات البارامترية لفروق المتوسط.
3- الاختبارات اللابارامترية للفروق النظّامية.
4- تقييم العلاقة بين المتغيرات.

أولاً: الإحصاء الوصفي:

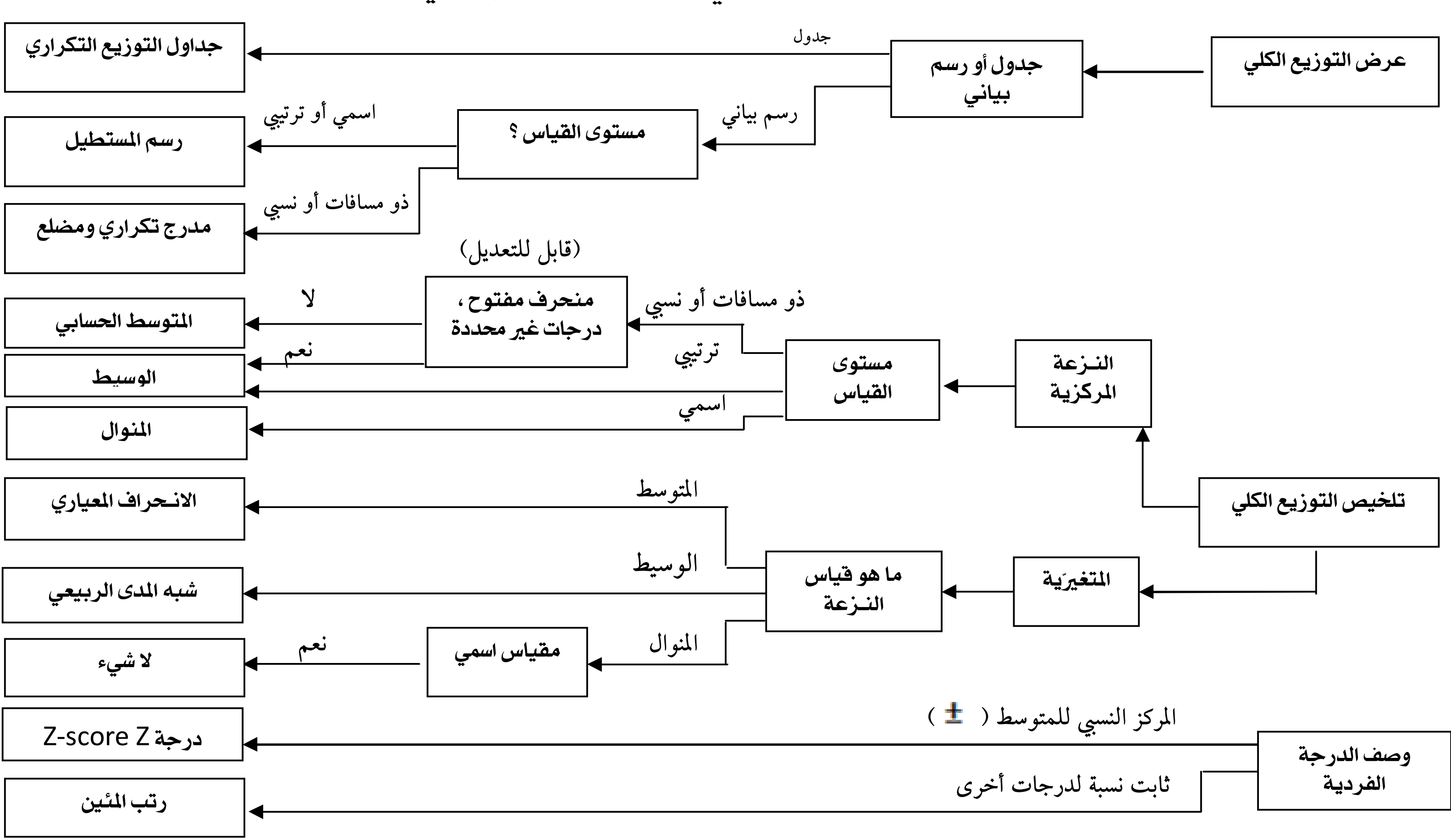
إن الهدف من وراء استخدام الأساليب الإحصائية الوصفية هـو تـسهيل وتنظـيم مجموعة الدرجات، فالدرجات يمكـن تنظيمهـا في جـداول إحـصائية أو رسـوم بيانيـة، أو يمكن تلخيصها من خلال حساب قيمة أو قيمتين تصفان مجموعة البيانـات بكاملـها، وأن أكثر الأساليب الإحصائية شيوعاً واستخداماً تتمثل في(*):

أ - جداول التوزيعات التكرارية أو الرسوم البيانية.
ب- مقاييس النـزعة المركزية.
ج- مقاييس التشتت (المتغيّرية).
د- درجات Z-Scores Z.

والشكل (1) خارطة قرار تساعد الباحث في اختيار الأسـاليب الإحـصائية الوصـفية لتحليل بياناته.

(*) يمكنك الرجوع إلى المسرد الذي يتضمن تعريفات هذه المفاهيم الإحصائية المرتبطة بالإحصاء الوصفي ملحق (2).

شكل (3-1): خارطة قرار تساعد الباحث في اختيار الإحصاء الوصفي لتحليل بياناته



ثانياً : الاختبارات البارامترية : تقييم فروق المتوسط بين مجتمعين :

كل الاختبارات التي يغطيها هذا الجزء تستخدم المتوسط الذي يتم الحصول عليه من بيانات عينة كأساس لاختبار الفروض حول متوسط المجتمع. وبالرغم من وجود عدة مواقف بحثية، إلا أنها في ذات الوقت تستخدم المنطق الأساسي نفسه، وأن كل الاختبارات الإحصائية لديها نفس البناء الأساسي، وفي كل حالة، فإن إحصاء الاختبار (Z، t أو F) تشتمل على حساب نسبة ratio تبعاً للبناء التالي:

$$\text{إحصاء الاختبار} = \frac{\text{الفرق المتحصل عليه بين المتوسطات}}{\text{فرق المتوسط المتوقع بالصدفة}}$$

إن الهدف من وراء كل اختبار هو تحديد ما إذا كانت فروق المتوسط المشاهد تكون أكبر من تلك الفروق بالصدفة. وبعبارات عامة، فإن نتيجة الدلالة تعني أن النتائج المتحصل عليها في دراسة بحثية (فروق المتوسط) هي أكثر مما يتوقع من الفروق التي جاءت عن طريق الصدفة. وفي كل حالة فإن القيمة الكبيرة لنسبة إحصاء الاختبار تشير إلى نتيجة دالة.

إن الحسابات الفعلية تختلف قليلاً عن إحصاء اختبار لآخر، إلا أن هذه الاختبارات في المحصلة النهائية تشتمل كلها على نفس الحسابات الأساسية.

1- مجموعة درجات (عينة) تم الحصول عليها من كل مجتمع أو من كل حالة معالجة.

2- يتم حساب المتوسط الحسابي لكل مجموعة درجات، وأن بعض مقاييس المتغيّرية SS (مجموع التربيعات، الانـحراف المعياري أو التباين) تم الحصول عليه لكل مجموعة فردية من الدرجات.

3- الفروق بين متوسط العينات يمدنا بقياس كم يوجد من فرق بين أحوال المعالجة. إن سبب فروق المتوسط هذه، يمكن أن تكون حالات المعالجة سبباً في هذه الفروق، وعادة ما يُطْلَق عليها "فروقاً نظاميةً أو فروقاً متنبأ بها". هذه الفروق هي بسط Numerator معادلة إحصاء الاختبار.

4- المتغيرّية داخل كل مجموعة من الدرجات تمدنا بقياس غير نظامي أو فروق غير متنبأ بها من الفروق راجعة بالأساس للصدفة. ولأن الأفراد داخل كل حالة معالجة قد تمت معالجته بشكل متساوٍ تماماً، فإنه لا يوجد أي شيء يمكن أن يكون سبباً في أن تكون درجاتهم مختلفة. وعليه، فإن أية فروقٍ مشاهدة (المتغيرّية) داخل المعالجات افترض بأنها راجعة إلى الخطأ أو الصدفة.

واضعين كل ذلك في الاعتبار، فإن نسب كل إحصاء اختبار يمكن وصفها كالتالي:

$$\text{إحصاء الاختبار} = \frac{\text{الفروق(المتغيرية) بين المعالجات}}{\text{الفروق(المتغيرية) داخل المعالجات}}$$

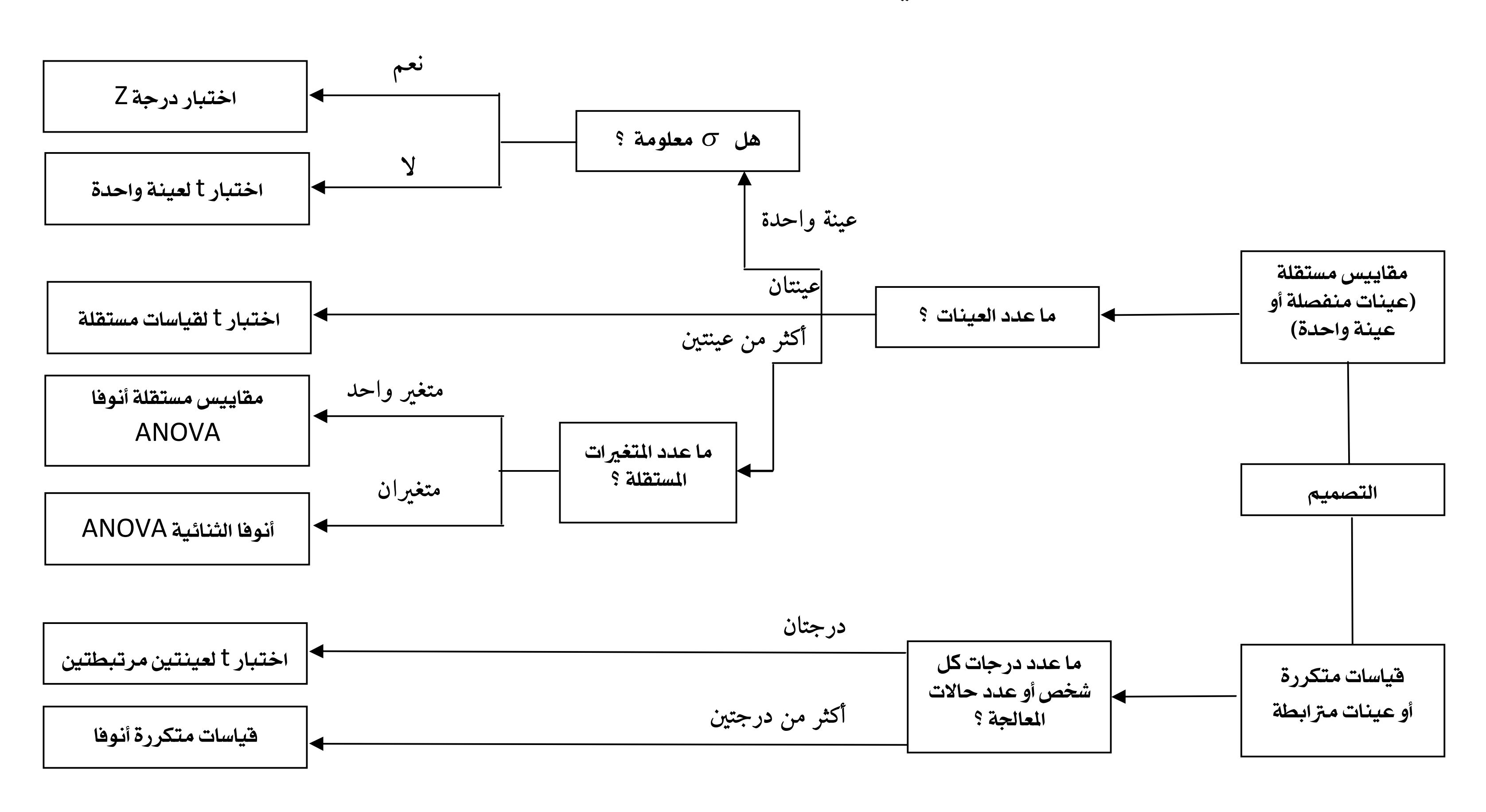
إن اختبارات الفروض التي تمت معاينتها في هذا الجزء ينطبق على ثلاثة تصميمات أساسية:

1- تصميمات لعينة مفردة (واحدة)، بيانات من عينة واحدة تستخدم لاختبار فرض حول مجتمع مفرد.

2- تصميمات مقاييس مستقلة: عينة مستقلة يتم الحصول عليها لتمثل كل مفردة أو حالة معالجة.

3- تصميمات لعينات مترابطة: إن أكثر التصميمات شيوعاً يعرف بِـ: تصميمات مقاييس متكررة التي تكون فيها نفس مجموعة الأشخاص يشاركون في كل حالات المعالجة المختلفة. وكذلك تتضمن هذه التصميمات، تصميمات الموضوعات المرتبطة التي تستخدم مجموعات منفصلة لكل حالة معالجة يتطلب فيها أن كل شخص في مجموعة واحدة يكون مرتبطاً بشكل متساوٍ مع الشخص الآخر في المجموعات الأخرى. والشكل (2) خارطة قرار تساعد الباحث في اختيار الاختبار البارامتري المناسب لتقييم فروق المتوسط بين مجتمعين.

شكل (3-2): خارطة قرار تساعد الباحث في اختيار الاختبار البارامتري المناسب لتقييم فروق المتوسط بين مجتمعين



وأخيراً، ينبغي علينا أن نكون حذرين بأن كل الاختبارات البارامترية تضع قيوداً صارمة على بيانات العينة وتوزيعات المجتمع الخاضع للدراسة. وأن أول هذه القيود، أن هذه الاختبارات تتطلب أن يكون مستوى القياس ذي المسافات أو النسبي (قيم عددية تسمح للباحث أن يقوم بحساب المتوسط، أو الفروق). وثاني هذه القيود تتمثل في أن يكون هناك افتراض لكل اختبار حول توزيع المجتمع وتقنيات المعاينة.

1- درجة Z Z- Score.
2- اختبار t لعينة واحدة.
3- اختبار t لمقاييس مستقلة.
4- اختبار t لعينتين مرتبطتين.
5- مقاييس مستقلة لأنوفا (أحادي الجانب).
6- مقاييس مستقلة لأنوفا (ثنائي الجانب).

ثالثاً: الاختبارات اللابارامترية: تقييم الفروق النظامية بين مجتمعين:

بالرغم من أن الاختبارات البارامترية هي أكثر شيوعاً واستخداماً للتقنيات الاستدلالية، إلا أن هناك عدة مواقف بحثية قد لا يكون بإمكان الباحث استخدام هذه الاختبارات البارامترية. وتندرج هذه المواقف ضمن ثلاث فئات عامة:

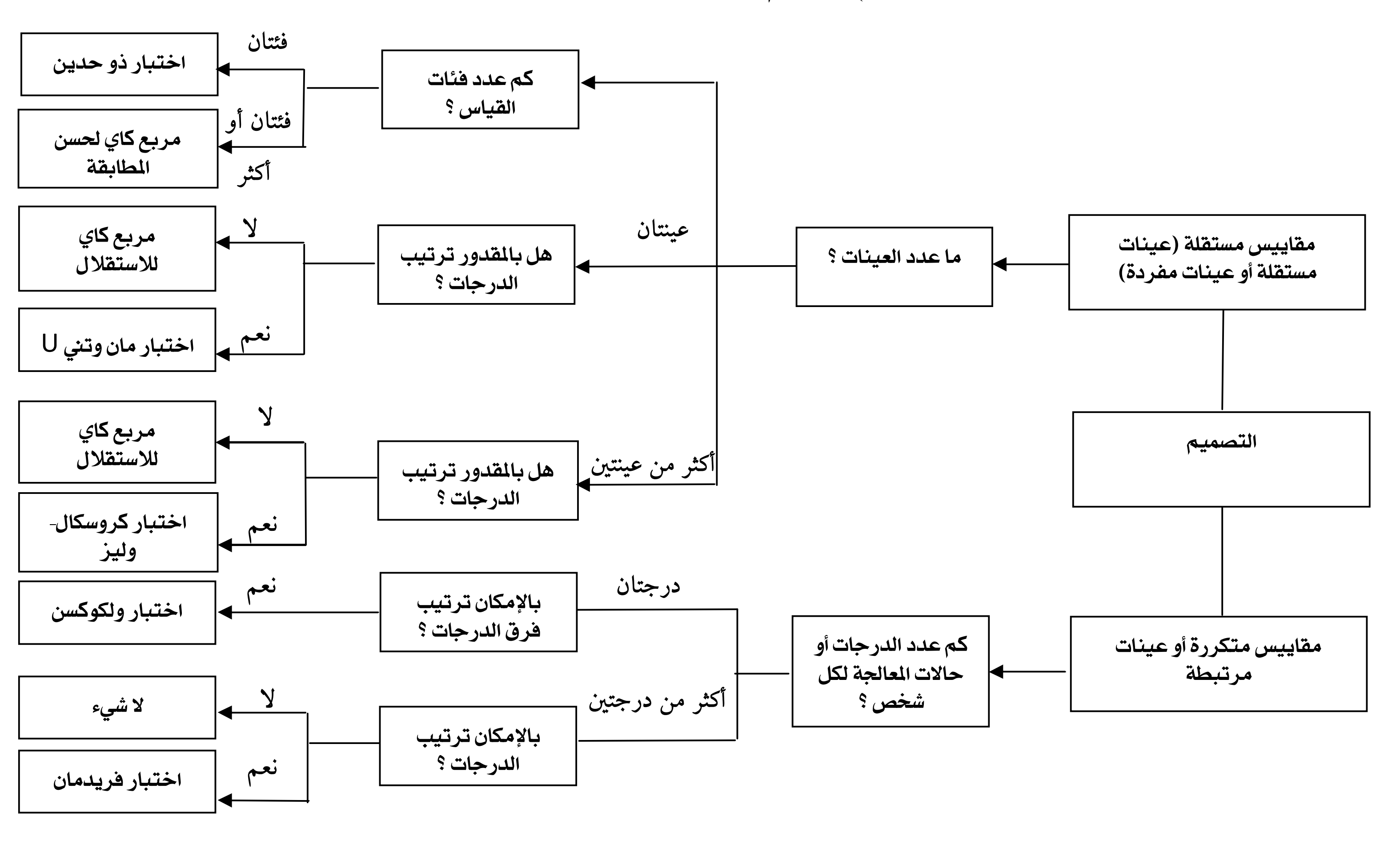
1- تتطلب البيانات مقاييس على المستوى الاسمي أو الترتيبي. وفي هذه المواقف يكون بمقدور الباحث حساب المتوسطات والتباين بإعتبار أن هذه القيم تمثل جزءاً أساسياً للاختبارات البارامترية.

2- عندما لا تَفّي بيانات هذه الاختبارات بالافتراضات التي تستند عليها الاختبارات البارامترية.

3- تنطوي البيانات على درجة قصوى من التباين التي تفوض احتمالية الدلالة للاختبار البارامتري. في هذه الحالة، ينبغي أن تحول الدرجات إلى فئات أو رتب، وبالتالي يكون الاختبار اللابارامتري هو البديل.

وعندما يكون الاختبار البارامتري لا يمكن استخدامه، وبالتالي تكون الاختبارات اللابارامترية هي البديل الملائم. وبشكل عام، فإن الاختبارات البارامترية تعتبر اختبارات أقوى بكثير إذا ما قورنت بالاختبارات اللابارامترية.

شكل (3-3): خارطة قرار تساعد الباحث في اختيار الاختبار اللابارامتري باستخدام الدرجات غير العددية (الرتب أو فئات اسمية) لتقييم فروق نظامية بين مجتمعين



ومهما يكن الأمر، فإنه عندما لا تكـون الاختبـارات البارامتريـة غـير ملائمـة، فـإن الاختبارات اللابارامترية تمد الباحثين بتقنيـة إحـصائية بديلـة لإجـراء التحليـل والتفـسير الإحصائي للنتائج البحثية. وتشتمل الإحصاءات اللابارامترية على:

1- اختبار مربع كاي لحسن المطابقة.
2- اختبار مربع كاي للاستقلال.
3- الاختبار ذي الحدين.
4- اختبار مان – وتني U.
5- اختبار ولكوكسن T.
6- اختبار كروسكال – وليز.
7- اختبار فريدمان.

رابعاً: مقاييس العلاقة بين المتغيرات:

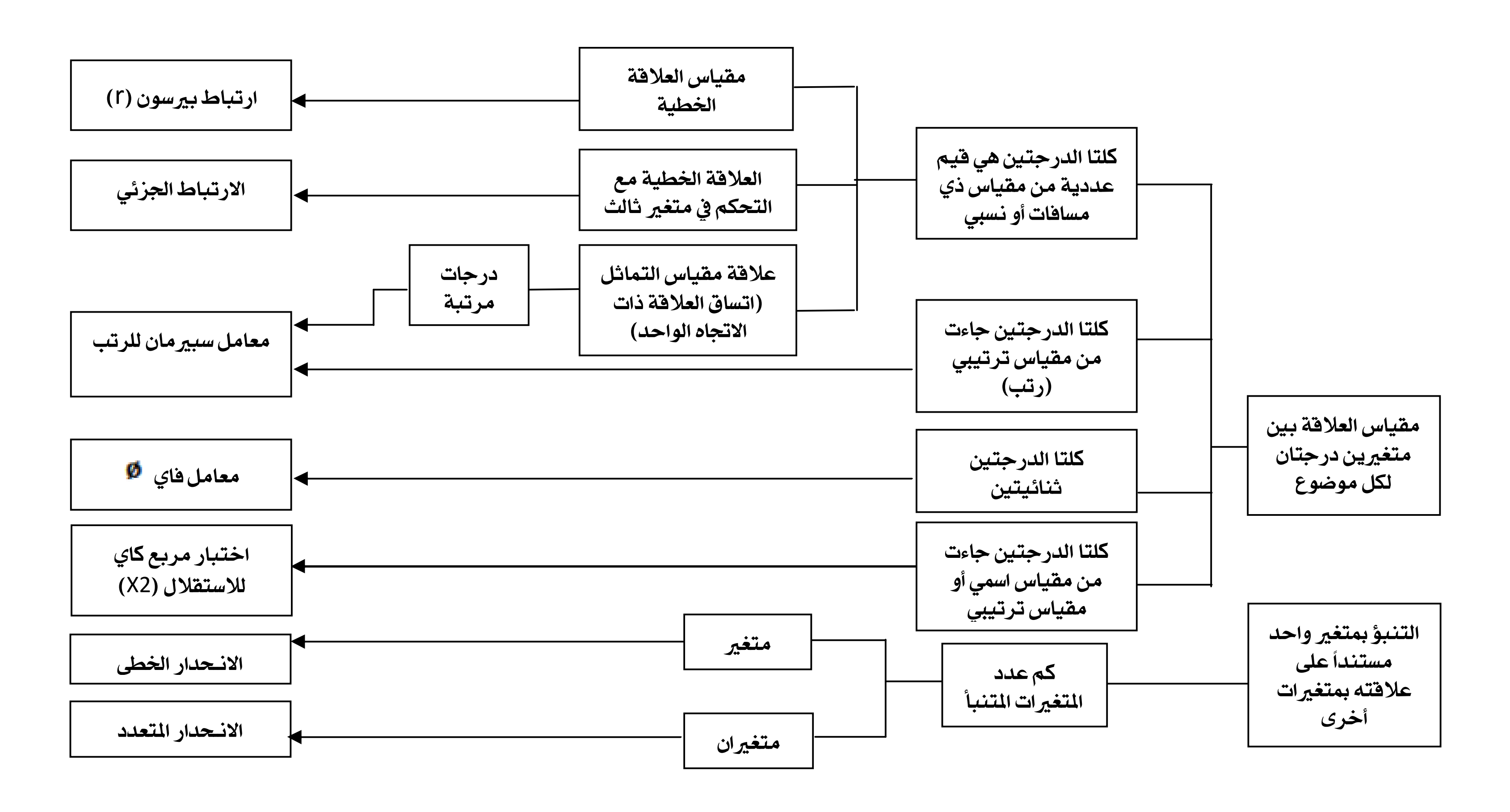
كل الطرق الإحصائية المختلفة التي تم وصفها في هـذا الجـزء يـراد منهـا أن تـستخدم نمطاً محدداً من البيانات. ولكي نـحدد الطريقة الملائمة، ينبغي على الباحث بادئ ذي بـدء أن يفحص بياناته ويعيّن ما هو نمط المتغير المستخدم، وما هـو مـستوى القيـاس المـستخدم لترميز المشاهدات.

وتشمل مقاييس العلاقة بين المتغيرات المقاييس التالية:

1- معامل بيرسون (r).
2- المعامل الجزئية $r_{yz.A}$.
3- معامل سبيرمان r_s.
4- معامل ϕ .
5- اختبار مربع كاي للاستقلال.
6- الانـحدار الخطي regression Line.
7- الانـحدار المتعدد Multiple Regression.

والشكل (4) خارطة قرار تساعد الباحث في اختيار الأسلوب الملائم لتقييم العلاقـة بين المتغيرات.

شكل (3-4): خارطة قرار تساعد الباحث في اختيار الأسلوب الملائم لتقييم العلاقة بين المتغيرات



ملحق (4)
بعض المعادلات الأساسية الواردة في الكتاب

1- متوسط مجتمع: بيانات مدرجة في قائمة: $u = \frac{\Sigma X}{N}$

N تشير إلى حجم المجتمع.

X تشير لكل درجة في التوزيع.

2- متوسط عينة: بيانات مدرجة في قائمة: = $\overline{X} = \frac{\Sigma X}{n}$

n تشير إلى حجم العينة.

3- متوسط عينة: بيانات تكرارية: $\overline{X} = \frac{\Sigma FX}{n}$

F تشير إلى التكرار لكل قيمة في التوزيع.

4- متوسط عينة: بيانات في فئات: $\overline{X} = \frac{\Sigma FM}{n}$

M تشير إلى مركز الفئة.

5- الانـحراف المعياري لمجتمع: بيانات مدرجة في قائمة: $\sigma = \sqrt{\frac{\Sigma(X-u)^2}{N}}$

6- الانـحراف المعياري لعينة: بيانات مدرجة في قائمة: $S = \sqrt{\frac{\Sigma(X-\overline{X})^2}{N-1}}$

$$S = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}{n-1}}$$

7- الانـحراف المعياري لعينة: بيانات تكرارية: $S = \sqrt{\frac{\Sigma FX^2 - \frac{(\Sigma FX)^2}{n}}{n-1}}$

8- معامل التباين النسبي (CRV): $CRV = \frac{S}{\overline{X}} \times 100$

9- مؤشر التباين النوعي (IQV): $IQV = \frac{\text{فروق مشاهدة}}{\text{أقصى فروق محتملة}}$

K تشير إلى عدد الفئات. $\frac{n^2(K-1)}{2K}$ = أقصى فروق محتملة

10- درجة Z لوصف مجتمع: $Z = \frac{X-\mu}{\sigma}$

11- درجة Z لوصف عينة: $Z = \frac{X-\overline{X}}{S}$

12- لامبيدا Lambeda $\lambda = \frac{E_1 - E_2}{E_1}$

E_1 تشير إلى عدد الأخطاء بدون معلومات حول المتغير المستقل.

E_2 تشير إلى عدد الأخطاء بمعلومات عن المتغير المستقل.

13- جرامرز $V = \sqrt{\frac{X^2}{n(K-1)}}$ Cramer`s V:

X^2 إحصاء مربع كاي لجدول التقاطع.

K عدد الصفوف أو عدد الأعمدة أيهما أصغر.

14- جاما GAMMA $G = \frac{N_C - N_d}{N_C + N_d}$

N_C تشير إلى عدد الأزواج المتوافقة.

N_d تشير إلى عدد الأزواج غير المتوافقة.

15- سومرز d: Somer`s d $d = \frac{N_C - N_d}{N_C + N_d + t_Y}$

t_y تشير إلى عدد الحالات المرتبطة على المتغير التابع مع التباين على المتغير المستقل.

16- كندل تاو – b: $tau-b=\frac{N_C-N_d}{\sqrt{(N_C+N_d+T_Y)(N_C+N_d+T_X)}}$

T_X تشير إلى عدد الحالات المرتبطة على المتغير المستقل مع التباين على المتغير التابع

17- كندل تاو - C: $tau-C=\frac{2K(N_C-N_d)}{N^2(K-1)}$

18- معامل سبيرمان للرتب: $r_S=1-\frac{6\Sigma D^2}{n(n^2-1)}$

19- معادلة الخط المستقيم. $Y=a\pm bX$

Y تشير إلى المتغير التابع.

X تشير إلى المتغير المستقل.

a كمية ثابتة وهي المساقة على محور (y) من البداية إلى النقطة التي تقطع محور (y) فهي قيمة (y) لقيمة (X) التي تساوي صفراً (0).

b تشير إلى ميل الخط.

+ تشير إلى التطابق الموجب.

– تشير إلى التطابق السالب.

20- معامل الانــحدار:

$$b=\frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$$

$$b=\frac{N\Sigma(xy)-(\Sigma x)(\Sigma y)}{N\Sigma x^2-(\Sigma x)^2}$$

21- معامل ارتباط بيرسون (r):

$$r=\frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma\left[(x-\bar{x})^2\right]\left[(y-\bar{y})^2\right]}}$$

$$r=\frac{N\Sigma(xy)-(\Sigma x)(\Sigma y)}{\sqrt{\left[N\Sigma x^2-(\Sigma x)^2\right]\left[N\Sigma y^2-(\Sigma y)^2\right]}}$$

22- فترات الثقة للمتوسط:

$$\text{Lower Limit الحد الأدنى} = \overline{X} - Z\left(\frac{S}{\sqrt{n}}\right)$$

$$\text{upper Limit الحد الأعلى} = \overline{X} + Z\left(\frac{S}{\sqrt{n}}\right)$$

23- اختبار Z لمتوسط مفرد: $Z = \frac{\overline{x} - u}{\sigma / \sqrt{N}}$

24- اختبار t لمتوسط عينة واحدة: $= \quad t = \frac{\overline{X} - u}{S / \sqrt{N}}$

25- اختبار Z لنسبة ذي حدين:

حيث إن $\quad Z = \frac{(P_S - 0.5) - P_u}{\sqrt{\frac{P_u(100 - P_u)}{n}}} \quad P_S > P_u$

أو

حيث إن: $\quad Z = \frac{(P_S + 0.5) - P_u}{\sqrt{\frac{P_u(100 - P_u)}{N}}} \quad Pu > Ps$

$$\sigma P = \sqrt{P_u\left(\frac{1 - P_u}{n}\right)}$$

Pu تشير إلى نسبة المجتمع.

26- اختبار مربع كاي لحسن المطابقة والاستقلال $x^2 = \sum \frac{(F_O - F_e)^2}{F_e}$

Fo تشير إلى التكرار المشاهد.

Fe تشير إلى التكرار المتوقع.

df = (r-1) (C-1)

حيث إن: r تشير إلى عدد الصفوف

C تشير إلى عدد الأعمدة

27- اختبار t لمتوسطين متساويين: $t = \frac{\overline{X}_1 - \overline{X}_2}{\sigma \overline{X} - \overline{X}}$

(تقدير التباين المجمع) $\sigma\overline{\mathbf{x}} - \overline{\mathbf{x}} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$

28- اختبار F لأكثر من متوسطين $F = \frac{\frac{SSB}{K-1}}{\frac{SSW}{n-K}}$

$$TSS = SSB + SSW$$

$$TSS = \sum X - m\overline{X}^2$$

$$SSW = \sum (X - \overline{X}_S)^2$$

$$SSB = \sum N_S (\overline{X}_S - \overline{X})^2$$

$\overline{X}_s$ تشير إلى متوسط لعينة معطاة.

n_s تشير إلى عدد الحالات لعينة معطاة.

29- اختبار t لفرق المتوسط لعينتين مرتبطتين $t = \frac{\overline{X}_D}{S_D / \sqrt{n}}$

$$\overline{X}_D = \frac{\Sigma D}{n}$$

$$S_D = \sqrt{\frac{\Sigma D^2 - \frac{(\Sigma D)^2}{n}}{N-1}}$$

30- اختبار مان – وتني U $UA + UB = n$

$$U_A = n_V(n_B) + \frac{n_A(n_A + 1)}{2} - \sum R_A$$

$$U_B = n_A(n_B) + \frac{n_B(n_B +_1)}{2} - \sum R_B$$

التقريب الطبيعي لاختبار مان) وتني U

$$Z = \frac{x-u}{\sigma} = \frac{\frac{U - n_A n_B}{2}}{\sqrt{\frac{n_A n_B (n_A + n_B + 1)}{12}}}$$

31- اختبار كروسكال – وليز $H = \frac{12}{N(N+1)}\left(\sum \frac{T^2}{n}\right) - 3(N+1)$

32- اختبار فريدمان $X_r^2 = \frac{12}{nk(k+1)}\sum R^2 - 3n(k+1)$

33- مقاييس حجم التأثير Cohen`s d = $\frac{\text{فرق المتوسط}}{\text{الانحراف المعياري}}$

(نسبة التباين التي تدخل في الاعتبار) r^2 n^2 and

(لاختبار t المستقلة والمتكررة) $r^2 =$ $r_2 = \frac{t_2}{t_2 + df}$ أو Eta^2

(لاختبار مربع كاي للاستقلال) $phi\sqrt{\frac{x^2}{n}}$

Cramer's $V\sqrt{\frac{x^2}{n(dF)}}$

المصادر

أولاً: المراجع العربية.

1- جولي بالانت (ت) خالد العامري، التحليل الإحصائي باستخدام SPSS، دار الفاروق للنشر والتوزيع، مصر، 2009م.

2- عبد الجبار توفيق البياتي، البحث التجريبي واختبار الفرضيات في العلوم التربوية والنفسية، جهينة للنشر والتوزيع، عمان، 2006م.

3- عبد الله عامر الهمالي، أسلوب البحث الاجتماعي وتقنياته، ط3، منشورات جامعة قاريونس، 2003 م.

4- ــــــــــــــ، التقنيات الإحصائية ومناهج البحث، منشورات جامعة قاريونس، 2008 م.

5- ــــــــــــــ، التحديث الاجتماعي معالمه ونماذج من تطبيقاته، منشورات جامعة قاريونس، 2008م.

6- سعد اللافي، الإحصاء الاستنتاجي، ط 1، منشورات أكاديمية الدراسات العليا، طرابلس، 2003م.

7- محمد صبحي أبو صالح وعدنان محمد عوض، مقدمة في الإحصاء، مكتب الكتب الأردني، عمان، الأردن، 1990م.

8- مصطفى خلف عبد الجواد، الإحصاء الاجتماعي: المبادئ والتطبيقات، دار المسيرة للنشر والتوزيع والطباعة، الأردن، 2009م.

ثانياً : المراجع باللغة الإنجليزية :

1. Colin Robson, Real World Research , Black well Publishing , USA , 2005.
2. EARL BABBie , The basic of Social Research , Thomson Wadsworth , USA , 2005.
3. Frederick J. Gravetter and Larry B. Wallnau , Statistics for The Behavioral Sciences , 8th ed , Wadsworth Cengage Learning , USA , 2010.
4. George Argyrous , Statistics for Social and Health Research With a Guide to SPSS , Sage Publications , London , 2001.
5. George DiEkhoff, Statictics for the social and Behavioral Sciences: Univariat, Bivariate, Multivariate , WM. C. Brown Publications , USA , 1992.
6. Hubert Blalock, Jr. Social Statistics , Mcgraw Hill Company , Inc. , New york , 1972.
7. Hugh Coolicam and Hodder Arnold , Research Methods and Statistics in Psychology , 4 th ed , Hodder and Stoughton , Educational , London , 2004.
8. J. Richard Kendrick , Social Statistics: An Introduction using SPSS for Windows , 2th ed , Pearson Education , Inc. , 2005.
9. Joseph F. Healey , Statistics: A Tool for Social Research , Wadsworth Cengage Learning , 7th ed , USA , 2005.
10. ______________ , The Essentials of Statistics: A Tool for Social Research , Wadsworth Cengage Learning , USA , 2010.
11. N. J. Cohen , Statistical Power analysis for behavioral Science , SAN Diego , CA. Accademic Press , 1988.
12. R. Mark Sirikin , Statistics for Social Science , Sage Publications , International Oaks , London , 1995.
13. William E. Wagner , III , 2th ed , using SPSS for Social Statistics and Research Methods , Pine force Press , an Inprint of Sage , Publications , Inc. , USA , 2010.



المؤلف : أ. د. عبد الله عامر الهمالي

المؤهلات العلمية :

- ليسانس علم الاجتماع والفلسفة ، 1970 ، كلية الآداب ، الجامعة الليبية.
- ماجستير علم الاجتماع ، 1974 ، الولايات المتحدة .
- دكتوراه علم الاجتماع ، 1979 ، الولايات المتحدة .

الخبرة الإدارية :

- رئيساً لقسم علم الاجتماع - جامعة بنغازي (قاريونس سابقاً)، 1985-1994م
- عميداً لكلية الآداب - جامعة بنغازي (قاريونس سابقاً) ، 1994 - 1997 م
- مديراً للإدارة العامة للعلاقات العلمية والتعاون الفني بقطاع التعليم والبحث العلمي 1998 م .
- مديراً عاماً للإدارة العامة للبعثات الدراسية والتدريب بقطاع التعليم والتكوين المهني 1999-2000 م .
- أميناً لمركز البحوث والدراسات 2003-2008 م .
- نائباً لرئيس منظمة التقدم الدولية (سابقاً) فيينا-النمسا .

المؤلفات والبحوث العلمية :

- المرأة العربية والمشاركة السياسية : دراسة ميدانية (بالاشتراك) 1983م
- التغير الاجتماعي : أسسه وتطبيقاته (بالاشتراك) 1984 م
- التحديث الاجتماعي : معالمه ونماذج من تطبيقاته 1986 م
- أسلوب البحث الاجتماعي وتقنياته 2003 م
- التقنيات الإحصائية ومناهج البحث 2008 م
- إضافة إلى مجموعة بحوث علمية في مجال تخصصه منشورة في كتب ومجلات علمية متخصصة

الخبرة التدريسية :

- أستاذ علم الاجتماع بجامعة بنغازي والجامعات الليبية وخبرة تدريسية تزيد عن ثلاثين سنة في مجالات : مناهج البحث الاجتماعي ، تصميم البحوث الاجتماعية ، الإحصاء الاجتماعي ، التغير الاجتماعي والتحديث .

الندوات والمؤتمرات والملتقيات العلمية :

- شارك في العديد من الندوات والمؤتمرات والملتقيات العلمية بالجامعات ومراكز البحث العلمي بلغت أكثر من خمس وعشرين مشاركة .
- شارك في عدة لجان علمية وفنية بلغت في مجموعها أربعون لجنة .
- أشرف وناقش أكثر من خمس وثلاثين رسالة علمية (ماجستير ودكتوراه) في الجامعات الليبية والعربية .

هذا الكتاب

يهدف هذا الكتاب المرجعي في المقام الأول إلى تعلم الأساليب الإحصائية واستخدامها في تحليل البيانات . كما يتيح أيضا فرصة تساعد الدارسين والمهتمين بمجال البحث العلمي بشكل عام، والبحث في مجال العلم الاجتماعي بشكل خاص في فهم الأساليب الإحصائية الوصفية والاستدلالية واستخداماتها الملائمة في تحليل البيانات ، ومنحهم الثقة اللازمة للتعامل مع هذه الأساليب الإحصائية واستخداماتها بشكل علمي دقيق . إضافة إلى هذا كله ، يتضمن هذا الكتاب بعض الإجراءات المتعلقة باستخدام المجموعة الإحصائية في العلوم الاجتماعية (SPSS) ، والكيفية التي تتم بها عملية تفسير مخرجات هذه المجموعة الإحصائية .

يتألف هذا الكتاب من عشرين فصلا تغطي بدءا من الإحصاءات الوصفية وحتى الإحصاءات الاستدلالية المتقدمة . وينقسم الكتاب في ذلك إلى ستة أجزاء . وينتهي بمسرد لأهم المفاهيم الإحصائية الوصفية والاستدلالية وتعريفها ، ومنظم إحصائي يلخص الأساليب الإحصائية التي تم تغطيتها في هذا الكتاب .